Journal of the
OPTICAL SOCIETY
of AMERICA

VOLUME 35, NUMBER 7 JULY, 1945

The Relationship Between the Granularity and Graininess of Developed Photographic Materials*

LOYD A. JONES AND GEORGE C. HIGGINS
Kodak Research Laboratories, Eastman Kodak Company, Rochester, New York

(Received May 21, 1945)

The objective quantity, "granularity," which refers to the spatial variations in transmitting or reflecting characteristics of a developed photographic material, has been determined for a group of materials differing widely in sensitivity and grain size. These measurements were made by several different methods which have been proposed by various workers. The psychophysical quantity, "graininess," which refers to the visual appearance of the granular structure in a developed photographic material, has been measured by the method proposed by Jones and Deisch. These measurements of graininess were made on the same samples as those which were used in the evaluation of granularity. None of the methods of measuring granularity give the same functional relation between granularity and the density of the silver image as that existing between graininess and density. However, by choosing arbitrarily a density level which is not the same for all of the objective methods, all these methods can be made to give granularity values which place the different photographic materials in *approximately* the same order as the graininess values, but for no method is the order *exactly* the same. Moreover, even when the order is the same, the granularity values are not proportional to the graininess values. These results indicate that the objective methods measure one or more, but not all, of the factors which determine graininess.

INTRODUCTION

DURING the past two or three decades much attention has been devoted to the construction of instruments and the development of techniques for measuring various aspects of the spatial inhomogeneities existing in developed silver images. The results of many of these investigations have been published in the literature.

One of the important aspects of these inhomogeneities is that which is apparent when the developed image is examined visually, either directly or by means of visual aids which result in magnification. The results obtained by the methods used for this purpose obviously depend to some extent upon the characteristics of the particular observer's visual mechanism and also upon the optical characteristics of the instruments employed. Many workers have sought to avoid the difficulties of the direct psychophysical evaluation of this image inhomogeneity by resorting to purely objective methods of measurement, and much effort has been directed toward the development of instruments and techniques which will give values in direct agreement with the visual appearance.

In the earlier publications dealing with these

* Communication No. 1035 from the Kodak Research Laboratories.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

problems, more attention was given to the psychophysical methods involving direct visual inspection than to the objective methods designed to express in correlated, but still purely objective, terms the spatial inhomogeneity of the developed silver deposits. The operational characteristics of the objective methods described in the literature are such that it is very difficult, if not impossible, to predict with certainty the correlation between the values obtained when these methods are applied to the same materials. Diverse and conflicting claims have been made by the proponents of these methods concerning the excellence with which the values represent the visual appearance of these silver deposits.

During more recent years we have given increasing attention to these objective methods and have constructed the variety of complex instruments necessary for the duplication of the work published by others. A series of carefully prepared samples has been measured by all of the methods which we believe warrant consideration. Since these measurements make it possible to establish the relations between the determinations based on different methods, and since, insofar as we are aware, no report has been published giving direct comparative results of measurements made on the same set of samples by the various suggested methods, our results may be of sufficient interest to workers in this field to warrant their publication. For the benefit of those readers who may have a general interest in the subject, but who may not have intimate knowledge of the published literature, a brief review of the literature and a summary of the present status of the investigational work in the field are given.

In a report published by one of us and Deisch[1] some years ago, the term "graininess" was defined as "the sensation or impression of nonuniformity in a photographic deposit produced on the consciousness of the observer when such a deposit is viewed." Thus, graininess was defined as a purely subjective concept and it was implied that the method described provided a means of measuring this subjective aspect of the inhomogeneity of a developed silver image. There have been voluminous discussions in the literature of psychology and psychophysics on the possibility of measuring a sensation or a perception, in the sense in which the word "measure" is used by physicists. Actually, the method suggested by Jones and Deisch is a psychophysical one and, while it cannot be maintained that it measures sensation or perception, it does measure or evaluate the purely physical or objective stimulus (the spatial inhomogeneity of the developed silver image) in terms of the subjective response. The relationship thus established is a psychophysical relationship.

In this communication we shall continue to use the word "graininess" when referring to the psychophysical aspect of the inhomogeneity of the developed silver image. It seems probable that in the near future it will be found necessary, or at least desirable, to modify and systematize the terminology used in referring to various aspects of the spatial structure of the developed photographic image. A course similar to that recently recommended by the Colorimetry Committee of the Optical Society of America[2] for dealing with terminology in the field of color seems promising. This would provide a definite term to be used as indicative of the physical (objective), psychophysical, and psychological (subjective) aspects of the image structure.* The term "granularity" will be used to designate the objective aspect evaluated in terms of spatial variations in the transmitting or reflecting properties of the developed photographic image. In the literature the term "graininess" has frequently been used for both the psychophysical quantity, graininess, and for the physical (objective) quantity, granularity. Both of these properties of the photographic material are important,

[1] L. A. Jones and N. Deisch, J. Frank. Inst. **190**, 657 (1920).

[2] Colorimetry Committee, J. Opt. Soc. Am. **34**, 183, 245 (1944).

* While we do not wish to make definite proposals at this moment, it seems logical to retain the word *graininess* when referring to the sensation or perception arising when a grainy visual field is observed. In the report of the OSA Colorimetry Committee it has been recommended that brightness and other words having the "-ness" ending be reserved for the various subjective responses arising in the consciousness of the observer as a result of objective stimuli. The word "luminance" was recommended as the psychophysical correlate of brightness. If a parallel course is followed in establishing a nomenclature for the granularity-graininess field, it would seem logical to use the terms *granularity*, *granulance*, and *graininess* for the *physical*, *psychophysical*, and *psychological* aspects of the structural characteristics of the developed photographic material.

but they are not identical and the two terms should not be used interchangeably.

REVIEW OF THE LITERATURE

Some of the earliest work done on the study of the graininess of photographic deposits was reported in publications from these Laboratories.[1,3] A method was used in which an enlarged image of the uniformly exposed and developed photographic material was moved away from the observer until it appeared homogeneous. The magnification was fixed and the relative merging distances for different samples were used to evaluate graininess. It was found possible to make measurements with a precision which gave an average deviation of ±3 percent. A similar method was used by Crabtree[4] to compare the graininess of motion-picture films. Lowry[5] studied graininess with an instrument in which the observer viewed an aerial image of the photographic deposit. The magnification of the virtual image was varied until the deposit appeared homogeneous. The relative magnifications required to give a homogeneous image with different materials were used to establish values of graininess. Conklin[6] studied graininess with a comparison microscope with which the samples were magnified and compared under the same viewing conditions. The granular structure of each sample was at all times visible, and the relative magnification of the two samples was adjusted until they appeared to be *equally inhomogeneous*. It is open to question whether the ratio of magnifications for two samples which made them appear equally inhomogeneous would always be the same as the ratio of the maximum magnifications at which the same samples would appear to be homogeneous.

Threadgold[7] suggested that graininess could be measured by the Callier coefficient Q (specular density/diffuse density). The use of the Q factor to measure graininess was studied by Küster,[8] and by Eggert and Küster.[9] They found a satisfactory agreement between graininess and the logarithm of the Q factor. Hansen and Keck[10] found a nearly linear relation between the logarithm of the Q factor and graininess as measured on an instrument similar to that of Jones and Deisch. Lowry, however, found a very poor correlation between the Q factor and graininess, and in one case he found the results contradictory. Eggert and Küster pointed out that "only the grains actually in the layer cause scattering of the light. Grain images printed from a negative onto a positive do not affect the light-scattering properties of that positive. If, for example, a negative having a Callier quotient of 2.0 is printed upon a grainless positive, the Callier quotient of that positive will still be 1.0, because there is no substantial graininess in the positive material. If, however, the print is examined, images of the negative grain will be seen." This means that the positive described by Eggert and Küster would appear non-homogeneous, or have graininess, when viewed by an observer, even though the Callier coefficient was 1.0.

Dunham[11] suggested that the granularity of a photographic material might be measured by the variations in a trace made with a microphotometer. Van Kreveld[12] and his co-workers developed a very ingenious method of measuring granularity by splitting into two parts the image of the sample formed by a microphotometer and measuring an average difference in transmittance between adjacent areas on a film. A method of measuring granularity based upon the standard deviation of the relative transmittance $\Delta T/T_m$ was developed by Goetz[13] and his co-workers. The quantity ΔT is the value of the individual transmittance fluctuations measured from the mean, and T_m is the mean transmittance. Goetz

[3] A. C. Hardy and L. A. Jones, Trans. Soc. Mot. Pict. Eng. No. 14, 107 (1922).
[4] J. I. Crabtree, Trans. Soc. Mot. Pict. Eng. No. 29, 77 (1927).
[5] E. M. Lowry, J. Opt. Soc. Am. **26**, 65 (1936).
[6] O. E. Conklin, J. Soc. Mot. Pict. Eng. **16**, 159 (1931).
[7] S. D. Threadgold, Phot. J. **72**, 348 (1932).
[8] A. Küster, Veröff. wiss. Agfa **3**, 93 (1933).
[9] J. Eggert and A. Küster, Kinotechnik **16**, 127, 291, 308 (1934); J. Soc. Mot. Pict. Eng. **30**, 181 (1938).
[10] G. Hansen and P. H. Keck, Zeits. f. wiss. Phot. **37**, 86, 99 (1938).
[11] T. Dunham, Jr., Ber. VIII internat. Kongr. Phot., Dresden, 287 (1931).
[12] A. van Kreveld, Phot. J. **74**, 590 (1934); J. Opt. Soc. Am. **26**, 170 (1936); A. van Kreveld and J. C. Scheffer, J. Opt. Soc. Am. **27**, 100 (1937).
[13] A. Goetz and W. O. Gould, J. Soc. Mot. Pict. Eng. **29**, 510 (1937); A. Goetz, W. O. Gould, and A. Dember, J. Soc. Mot. Pict. Eng. **34**, 279 (1940); A. Goetz and F. W. Brown, J. Soc. Mot. Pict. Eng. **39**, 375 (1942).

measured the relative transmittance fluctuations $\Delta T/T_m$ in order to approximate the logarithmic relation between stimulus and sensation. He suggested that since the logarithmic relationship holds only within certain ranges of illumination, the measurements in terms of $\Delta T/T_m$ must be corrected by the function for the change in contrast sensitivity of the eye with field brightness. However, it cannot be assumed without experimental evidence that there is a significant correlation between graininess and the values of granularity measured by his method. Goetz also suggested that the absolute variation in transmittance is important in many cases, such as sound reproduction, and pointed out that the granularity values obtained as a function of relative transmittance fluctuations $\Delta T/T_m$ may be changed to absolute values by multiplying by the mean transmittance T_m.

Selwyn[14] developed a theory of graininess based upon the statistical distribution of density measured over a large number of small areas on the developed photographic emulsion. Granularity was defined by Selwyn as the product of the standard deviation (root-mean-square) of the density from the mean and the square root of twice the scanning area.

The standard deviation of density or transmittance measured by a microphotometer depends upon the scanning area used. Van Kreveld and Goetz stated that the effect of the scanning area can be compensated for, since theoretically the standard deviation should vary inversely with the square root of the scanning area. While van Kreveld and his co-workers reported experimental evidence supporting the theory that the deviation varies inversely with the square root of the scanning area, Hansen and Keck, and Selwyn reported experimental evidence showing slight variations from this theory.

Romer and Selwyn[15] designed an instrument to measure granularity by the power of the photographic deposit to obscure fine detail. An enlarged image of the deposit being examined was thrown on a test chart consisting of twenty-six circular patterns, each containing ten pairs of lines. The spacing of the parallel lines in that test pattern for which only half of the pairs of lines could be resolved was found to be proportional to the value of granularity as measured in terms of standard deviation of density. All measurements were made in terms of a standard sample measured on a microphotometer. Selwyn[16] suggested that the graininess of a negative, for pictorial purposes, should be the quotient of the granularity, measured as described, and the slope of the characteristic curve at a point one log exposure unit above the point where the gradient of its $D-\log E$ curve is $0.3\bar{G}$, $\bar{G}$ being the average gradient between these two points.

EXPERIMENTAL

Nine photographic materials were used in this investigation. Some were included which are usually designated as negative materials, and others normally classified as of the positive type. For the purposes of this paper the distinction between positive and negative materials is meaningless except that, in general, negative materials have relatively coarse grain structures compared to positive materials.

The nine materials were selected to cover the entire range of graininess and granularity normally found in commercially available photographic materials. For convenience they have been designated by the numbers 1 to 9, inclusive, as shown in the first column of Table I. In the next column are given brief descriptions indicating in a general way their characteristics. In column 3 are given values of relative speed.* While there is no unique relation between the speed of a material and its graininess or granularity, it may be said that graininess and granularity usually tend to increase with increasing speed. The speed values shown in Table I are based on the fractional gradient criterion adopted by the American Standards Association as a standard method of expressing this characteristic of a limited group of photographic materials.[17]

[14] E. W. H. Selwyn, Phot. J. **75**, 571 (1935); **79**, 513 (1939).

[15] W. Romer and E. W. H. Selwyn, Phot. J. **83**, 17 (1943).

[16] E. W. H. Selwyn, Phot. J. **83**, 227 (1943).

* No value of speed is given for the x-ray material. The "white-light" speed of this material has no significance even as an approximate indication of its graininess or granularity when exposed to x-rays.

[17] American Standard Method for Determining Photographic Speed and Speed Number, ASA Z38.2.1–1943, J. Opt. Soc. Am. **33**, 479 (1943).

TABLE I. Description of the photographic materials and the development conditions used.

No.	Type of material	Rel. speed	Developer	Time in min. at 68°F	Gamma
1	X-ray	—	Kodalk x-ray	6	—
2	Fast panchromatic negative	1000	D-19	8	1.38
3	Medium-speed panchromatic negative	500	D-76	9	0.65
4	Fast blue-sensitive positive	50	D-16	5	1.39
5	Slow panchromatic negative	250	D-76	12	0.94
6	Slow blue-sensitive positive	10	D-16	4.5	2.08
7	Slow panchromatic positive	20	DK-50	5	2.80
8	Very slow panchromatic negative	10	D-76	6	0.75
9	Very slow, high contrast orthochromatic	5	D-85	2	6.20

We wish to emphasize that the present investigation is concerned only with the evaluation, in terms of graininess and granularity, of the structural characteristics of these developed photographic materials. They are regarded simply as examples of various granular structures and no consideration is given to the possible effects which these granular structures may exert on either the graininess or granularity of prints made therefrom.

All of these materials, with the exception of No. 1, were exposed in a time-scale sensitometer equipped with a tungsten light source having a color temperature of 2850°K. The x-ray material was exposed to 1,000,000-volt x-rays. X-rays were used instead of white light to obtain higher values of graininess and granularity than are secured on this same material with white-light exposures. It should be mentioned also that, although this material is normally coated with emulsion on both sides of the support, one of these coatings was removed prior to reading the graininess and granularity values in order to reduce the density sufficiently so that measurements could be made.

After exposure, the sensitometric strips were developed in a developing machine which is used for sensitometric work. The compositions of the developing solutions used are indicated in column 4 of Table I, and the development times were as shown in column 5. These development conditions are the same as those recommended by the manufacturer of the product and are representative of the treatment which these materials receive in practice.

Great care was exercised in preparing the sample sensitometric strips to insure that they were as free as possible from physical imperfections, such as scratches, pinholes, dust specks, abrasions, etc. Since samples are subjected to relatively high magnifications when granularity and graininess measurements are being made, the presence of any appreciable physical imperfections may give rise to serious errors of the values obtained. We cannot emphasize too strongly that the greatest possible care should be taken to avoid such sources of error.

GRAININESS MEASUREMENTS WITH THE JONES-DEISCH APPARATUS

The diagram of the Jones-Deisch instrument for measuring graininess is shown in Fig. 1. Light from the ribbon-filament lamp S_1 passes through the condenser C_1 and onto the sample O. The microscope K forms a magnified image of the sample on the magnesium carbonate screen Sc. The observer sees the image on this screen as reflected by the traveling mirror M. This mirror moves along the track T and is oriented by the template L, which is made so that the observer, looking through the artificial pupil E, always sees the image of the sample on the screen. Measurements were made by moving the mirror back until the image of the sample appeared homogeneous and then reading the blending distance on the tape at B. Each sample was read on at least two different days, and the maximum percentage variation between any two days was 10 percent; in most cases, it was less than 5 percent. The sample was magnified 34× on the screen, which had a field brightness, with no sample in place, of 15 foot-lamberts. In order to compensate for the changes in the visual acuity of the observer, the blending distance for a 500-line halftone screen was measured before and after each set of readings on the photographic material. The graininess data were computed by the following formula:

$$\text{Graininess} = 508D/d,$$

in which D = the merging distance for the sample, d = the merging distance for the 500-line screen, and $508 = 100{,}000/n$, where n is 197, the number of lines per cm for the halftone screen.

The values obtained by this method are listed in Table II, column C.

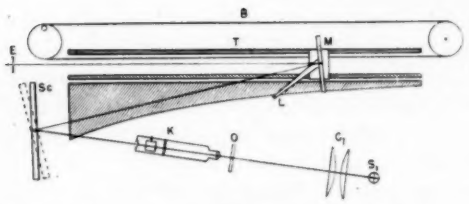

FIG. 1. Diagram of Jones-Deisch graininess instrument.

GRANULARITY MEASUREMENTS WITH THE VAN KREVELD TYPE OF APPARATUS

The schematic diagram for the van Kreveld type of apparatus used here is shown in Fig. 2. The light from a ribbon-filament lamp A passes through a neutral-density wedge B, and the filament is imaged on the sample D by an 8-mm microscope objective C. An image of the sample is formed on the double slit F by another 8-mm microscope objective E. Each slit was 4 mm wide by 7.5 mm long, which corresponded to an area of $6683\mu^2$ on the sample. A No. 920 twin-photo-tube G is placed in an enclosed box behind the slits in such a manner that the light passing through each slit falls on the cathode behind that aperture alone. The current from each unit of the photo-tube is amplified by twin d.c. amplifiers H. The output from one of these amplifiers is connected in such a way as to oppose that from the other, and as the sample moves past the objective, an alternating current is set up in the output circuit as the illumination on first one slit and then the other is the larger. The alternating current is then rectified by a barrier-layer rectifier I and fed into a ballistic galvanometer J, having a period of 20 seconds and a critical damping resistance of 17,000 ohms. The amplifier has a modified Miller[18] balanced cathode circuit.

The granularity measurements were made by moving the sample past the objective and reading the deflection of the galvanometer. The galvanometer reached equilibrium rapidly, and its deflection was a measure of an average difference in transmittance between the two areas scanned on the deposit. All readings were taken with constant illumination, each reading was repeated on several days, and the results were corrected for the nonlinearity of the rectifier for small currents.

[18] S. E. Miller, Electronics **14**, 27 (Nov. 1941).

The values obtained by this method are shown in Table II, column D.

GRANULARITY MEASUREMENTS WITH THE GOETZ-GOULD TYPE OF APPARATUS

Goetz and Gould defined a graininess coefficient by the distribution function,

$$y=(2/G'\sqrt{\pi})\int_0^x \exp\left[-(x/G')^2\right]dx, \qquad (1)$$

for the relative transmittance fluctuations in a photographic deposit. The graininess coefficient G, which is actually granularity, is equal to $1000G'$. The relative transmittance fluctuation x equals $\Delta T/T_m$, where ΔT is the value of the individual fluctuation from the mean transmittance T_m. The granularity coefficient G' is the standard deviation, σ, of the relative transmittance fluctuations from the mean transmittance multiplied by the square root of two ($G'=\sigma\sqrt{2}$).

The apparatus used to measure granularity as defined by Goetz and Gould consisted of two separate pieces of equipment, a microphotometer and an integrator. The regular laboratory microphotometer, equipped with a special amplifier and a recording lamp, was used to make all granularity traces. A schematic diagram of the microphotometer optics is shown in Fig. 3. The filament of the ribbon-filament lamp A is imaged on the auxiliary aperture C by the lens B. The auxiliary aperture C is imaged on the sample E by the lens D. The objective F images the sample on the main aperture G, while the condenser lens H images the objective F on the photo-tube I. With this optical system, the same area on the cathode of the photo-tube is illuminated at all times, irrespective of the size or position of the main aperture G.

The output of the photo-tube is fed into a linear d.c. amplifier. The output of the amplifier is fed into the galvanometer used to make the

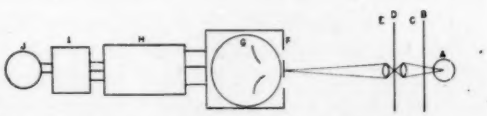

FIG. 2. Diagram for the van Kreveld type of apparatus for measuring granularity.

TABLE II. Granularity and graininess values.

No. A	Density* B	Jones-Deisch C	van Kreveld D	Goetz-Gould E	Goetz-Gould F	Selwyn G	Romer-Selwyn H	Küster-Eggert I	Küster-Eggert J	Küster-Eggert K
1	0.22	725	—	58	57	1.65	—	1.80	2.07	31.6
	.29	757	—	—	—	—	—	1.77	2.00	30.1
	.36	737	—	77	74	2.75	—	1.88	2.10	32.2
	.45	734	—	—	—	—	—	1.89	2.06	31.4
	.57	692	—	135	127	4.45	—	1.85	1.98	29.7
	.70	604	—	196	180	5.35	—	1.88	1.99	29.9
	.89	592	—	185	162	5.65	—	1.90	2.01	30.3
	1.01	492	—	197	165	5.15	—	—	—	—
	1.11	463	—	197	160	3.50	—	—	—	—
	1.23	447	—	—	—	3.40	—	—	—	—
2	.20	391	530	—	—	.90	—	2.20	2.57	40.0
	.31	455	490	39	34	1.15	—	2.14	2.34	36.9
	.47	432	430	54	52	1.25	.93	2.14	2.27	35.6
	.64	412	370	56	53	1.75	1.04	2.12	2.21	34.4
	.80	377	280	55	50	1.45	—	—	—	—
	1.03	311	210	58	50	1.45	—	—	—	—
	1.22	264	175	—	—	2.30	—	—	—	—
	1.39	242	140	—	—	1.75	—	—	—	—
	1.59	211	—	—	—	2.10	—	—	—	—
	1.76	—	—	—	—	2.55	—	—	—	—
3	.02	—	—	17	17	.45	—	—	—	—
	.05	190	380	19	19	—	—	—	—	—
	.08	217	415	20	20	.50	—	1.33	1.98	29.7
	.17	266	420	23	22	.55	.57	1.46	2.14	33.0
	.26	236	410	—	—	.90	.57	1.49	1.96	29.2
	.37	223	405	29	27	.80	.63	1.50	1.84	26.5
	.46	207	360	—	—	.75	.66	1.54	1.84	26.5
	.52	206	280	30	28	.75	.65	1.68	2.02	30.5
	.57	180	230	31	28	1.15	.67	1.54	1.74	24.1
	.72	175	220	29	26	1.15	.77	—	—	—
	.84	164	—	32	27	1.05	.80	—	—	—
	.90	160	—	31	26	1.35	.88	—	—	—
	.99	155	—	30	24	1.15	.91	—	—	—
4	.11	166	310	17	17	.25	—	1.77	1.87	27.2
	.17	217	350	19	19	.45	—	1.70	1.75	24.3
	.24	230	355	20	20	.75	.59	1.67	1.70	23.0
	.36	242	375	25	24	.65	.70	1.74	1.76	24.6
	.53	264	320	34	32	1.15	.75	1.69	1.71	23.3
	.68	258	260	34	32	.75	.83	1.79	1.80	25.5
	.90	232	200	32	29	1.10	.91	1.69	1.70	23.0
	1.12	230	120	—	—	1.35	.96	—	—	—
	1.33	184	—	—	—	1.50	—	—	—	—
5	0.09	184	280	18	17	0.20	—	—	—	—
	.15	201	360	21	20	.30	—	1.37	1.82	26.0
	.24	180	330	22	21	.65	0.54	1.46	1.90	27.9
	.37	166	290	23	22	.55	.63	1.53	1.89	27.6
	.45	165	280	26	24	.65	.70	1.53	1.81	25.8
	.53	165	250	26	24	.65	.79	1.51	1.70	23.0
	.68	164	220	29	26	.85	.81	1.47	1.61	20.7
	.86	160	160	31	27	1.00	.87	—	—	—
	.96	141	120	29	24	1.00	.94	—	—	—
6	.07	105	175	—	—	.20	—	1.42	1.55	19.0
	.17	115	210	13	13	.30	—	1.47	1.56	19.3
	.29	143	230	13	13	.20	.36	1.49	1.55	19.0
	.43	139	210	16	16	.35	.41	1.46	1.49	17.3
	.69	133	180	20	19	.25	.47	1.47	1.49	17.3
	.97	127	120	18	16	.45	.50	1.44	1.45	16.1
	1.22	—	—	—	—	.55	—	1.45	1.45	16.1
7	.16	90	180	—	—	.20	—	1.13	1.47	16.7
	.24	110	200	9	9	.25	—	1.25	1.60	20.4
	.40	125	190	13	12	.40	.45	1.22	1.38	14.0
	.67	125	140	15	14	.40	.51	1.31	1.45	16.1
	1.00	121	—	19	15	.50	.58	1.38	1.50	17.6
8	.09	96	170	—	—	.20	—	—	—	—
	.15	111	175	10	10	.30	—	1.17	1.54	18.8
	.27	109	180	13	12	.35	—	1.31	1.77	24.8
	.36	103	210	13	12	.50	—	1.22	1.40	14.6
	.45	96	170	14	13	.45	—	1.22	1.34	12.7
	.56	90	130	17	16	.45	—	1.25	1.38	14.0
	.69	84	125	14	12	.45	—	1.27	1.38	14.0
	.82	—	115	14	12	.65	.43	1.26	1.36	13.4
	.88	—	—	14	12	.55	.43	1.22	1.29	11.1
9	.15	90	190	13	13	.30	—	1.33	1.27	10.4
	.52	109	195	15	14	.65	.37	1.34	1.33	12.4
	1.25	106	120	19	15	.65	.42	1.28	1.28	10.7

* Diffuse density including fog, but not base.

microphotometer traces. The regular target lens used for the recording camera on the microphotometer was replaced by a lens one-half of which was silvered. The arrangement of the optical system is sketched in Fig. 4. The ribbon-filament lamp H is imaged on the target lens F by the lens G; this lens G is imaged on the lens E by the target lens F; the target lens, in turn, is imaged on the slit C of the recording camera. The cylindrical lens B, with its axis parallel to the camera slit, forms a line image of the target lens on the film A. The film A moves past the slit with a speed which is adjustable independently of the speed with which the sample is moved.

The trace on the recording film consists of a boundary with the area above clear and the area below blackened, as shown in Fig. 5. The abscissa on the trace in this figure represents linear distance on the sample. The linear magnification used was 200×, that is, one mm on the trace equals 1/200 mm on the sample. The ordinate on the trace represents the transmittance of the sample, and, as pointed out by Goetz and Gould, if the intensity of the light is regulated for emulsions of different densities so that the mean deflection of the recording galvanometer is constant, the ordinate is proportional to the relative transmittance. The sensitivity of the amplifier was held constant, and the light intensity was adjusted by means of a neutral wedge and small changes in lamp current to give the same mean deflection on the recording galvanometer for each sample. The aperture used to limit the scanning beam was a slit (G, Fig. 3) 40 mm long and 0.1 mm wide, which represented an area of 3.04 $\times 10^3 \mu^2$ on the sample.

A trace at least one meter long, with the sample moving 1 mm per minute, was made for each density step of the material being measured. A section 35 cm long was selected from the trace and analyzed in the integrator. The schematic

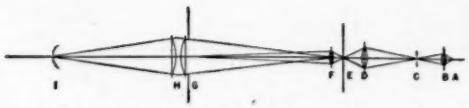

FIG. 3. Optical system of the microphotometer.

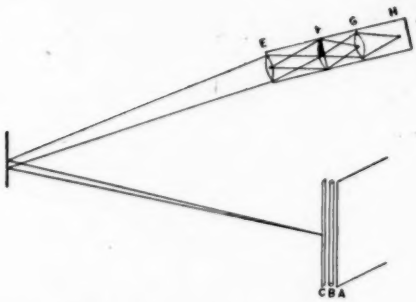

FIG. 4. Optical system for the target lens and recording camera on the microphotometer.

diagram of the integrator is shown in Fig. 6. The trace is illuminated evenly by the lamps behind the diffusing screen A and the trace B is scanned by a slit C, 0.15 mm wide, which is parallel to the abscissa of the trace. An image of the slit C is formed on the lens E by the objective D. The lens E images the objective D on the cathode of the photo-tube F. Two optical systems, both having a photo-tube, are used in order that a longer trace may be evaluated than would be possible with a single system. As the slit passes in front of the trace, the illumination on the photo-tube varies, but the same area on the cathode is illuminated at all times. The output from the photo-tubes is fed into the amplifier G, which is the same as that used in the microphotometer. The output from the amplifier is fed into the galvanometer H. The image of a point light source I is formed on the integrator recording paper J by the galvanometer mirror and lens. The frame holding the recording paper J is connected mechanically to the slit C so that, as the slit moves down across the trace, the recording paper moves up twice as fast.

The records were made by marking the path taken by the spot of light as it traveled across the paper when the slit was moved across the trace. The path was marked manually by dots on the paper and then checked by moving the slit across the trace again. The integrator records for two different microphotometer traces on the same sample are shown in Fig. 7. The abscissas on the record represent the displacements of the slit with reference to the ordinates on the trace. These ordinates on the trace represent the deflections of the recording galvanometer on the microphotometer, and, as pointed out before, these deflections are proportional to the relative transmittance of the original sample. The ordinates on the record represent the deflection of the galvanometer in the integrator; these deflections are a measure of the percentage of the total number of fluctuations on the trace having a given value of relative transmittance. The minimum ordinate, 0 percent, represents the deflection when the slit is over that area on the trace which is completely blackened; the abscissa corresponding to this minimum ordinate represents the maximum relative transmittance encountered in the original sample. The maximum ordinate, 100 percent, represents the deflection when the slit is over that area on the trace which is clear; the abscissa corresponding to this maximum ordinate represents the minimum relative transmittance encountered in the original sample. The value of the x coordinate on the integrator record corresponding to the y coordinate of 50 percent represents the mean value of relative transmittance. The deviations of the x coordinates on the integrator record from this mean value are a measure of the relative transmittance fluctuations $\Delta T/T_m$.

The curve in Fig. 7 is known as an ogive, and, if the distribution is Gaussian, it can be represented mathematically by Eq. (1), which is a

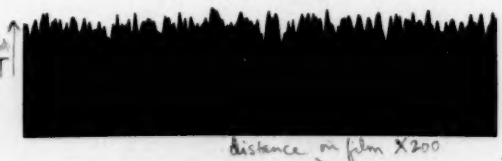

FIG. 5. A microphotometer record.

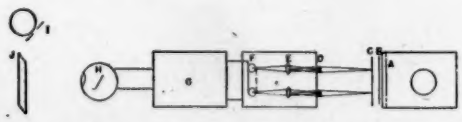

FIG. 6. Diagram of the integrator used to analyze the microphotometer record.

form of the Gaussian probability function. The value of G' in Eq. (1) can be obtained from the ogive by introducing a new y coordinate. Let $\theta(y)=x/G'$ and $Y=1/\theta(y)$, whence $G'=xY$. The product of the x and Y coordinates of any point on the ogive is the value of G'. The Y scales used to determine the value of G' were those shown in Fig. 8B of Goetz and Gould's first paper.[13] The granularity coefficient G, defined by Goetz and Gould as $1000G'$, was computed from the average of six different xY values on the integrator record.

The values by the Goetz-Gould method in Table II, column E, were measured in terms of the standard deviation of the relative transmittance fluctuations as described here. The granularity values in column F are the values in column E corrected for the change in contrast sensitivity of the eye. The curve used to make this correction is shown in Fig. 8.* The contrast sensitivity, or discrimination factor, $B/\Delta B$, is plotted as the ordinate and the density of the sample is plotted as the abscissa. For constant illumination, the field brightness varies with density, and for any level of illumination, the field brightness and the discrimination factor can be determined for each density. The values for what Goetz and Gould called "visual graininess," column F, are the product of granularity, column E, and the discrimination factor for that density as given by the curve in Fig. 8. The variation with density of the three types of granularity described by Goetz and Gould, for medium-speed panchromatic negative material No. 3, is shown in Fig. 9. The granularity G_a, which is a function of the absolute transmittance fluctuations, is the product of granularity for the relative transmittance fluctuations, column E, and the mean transmittance for that deposit.

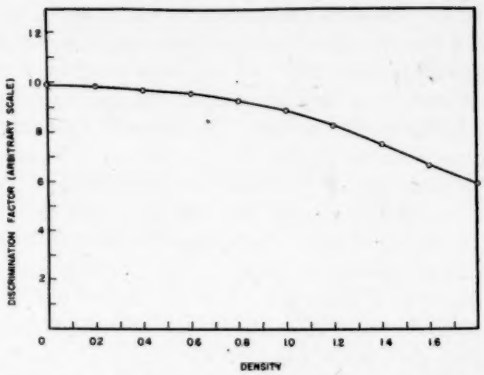

FIG. 8. Curve used to make correction for the change in the contrast sensitivity of the eye with brightness, as required by the Goetz-Gould method.

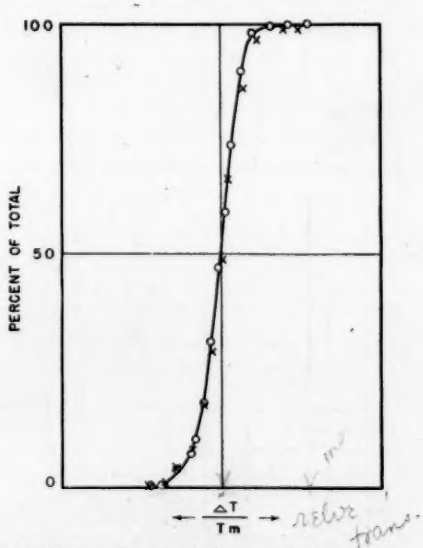

FIG. 7. Integrator records for two different microphotometer traces on the same sample.

* This curve is the same as that used by Goetz and Gould in applying the discrimination factor correction and is derived from the data published by Nutting (J. Opt. Soc. Am. 4, 55 (1920)).

GRANULARITY MEASUREMENTS BY THE SELWYN METHOD

Selwyn defined a granularity coefficient by the distribution function

$$y=(a/G^2\pi)^{\frac{1}{2}}\int_{-\infty}^{x}\exp\left(-x^2a/G^2\right)dx, \quad (2)$$

where x is the deviation of density from the mean, a is the scanning area, and G is the granularity coefficient. The granularity coefficient as defined by this equation is equal to the standard deviation of density from the mean multiplied by the square root of twice the scanning area, $[G=\sigma(2a)^{\frac{1}{2}}]$. Selwyn suggested that the values of this coefficient should be given in density-microns.

The apparatus used was the same as that for

the Goetz-Gould granularity measurements, except that the linear amplifier was replaced by a logarithmic amplifier and the main aperture was replaced by a circular aperture having an area of $1250\mu^2$ on the sample. The microdensitometer traces were similar to the trace shown in Fig. 5. The abscissas of the trace represent linear distances on the sample, and the ordinates represent the densities of the sample for the areas of the deposit being scanned. The traces were made with the illumination held constant for all densities, and they were integrated as described under the Goetz-Gould method. These integrator records are similar to that shown in Fig. 7, except that the abscissa now represents density fluctuations. The integrator data were then plotted on probability paper. This paper has a distorted ordinate scale so that, if the distribution is Gaussian, the ogive will plot as a straight line whose slope is the reciprocal of the standard deviation of the fluctuations. An example of such a plot of an integrator record is shown in Fig. 10. The granularity values obtained by the Selwyn method were determined by plotting all data on probability paper, measuring the slope to obtain the standard deviation, and multiplying the standard deviation so obtained by the square root of twice the scanning area (in microns). The values obtained by this method are shown in Table II, column *G*.

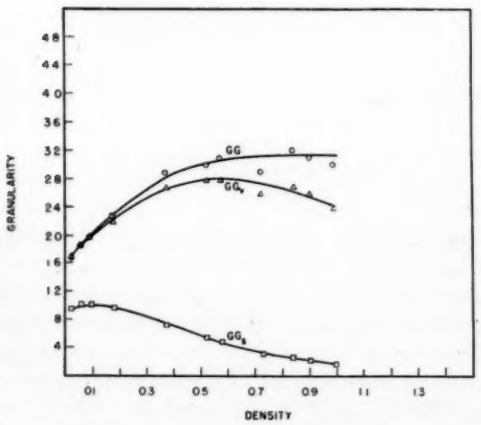

FIG. 9. The variation with density of the three types of granularity described by Goetz and Gould for medium-speed panchromatic negative material No. 3.

GRANULARITY MEASUREMENTS BY THE ROMER-SELWYN METHOD

A microscope was adapted to read granularity by the obscuring-power method described by Romer and Selwyn. The image of the deposit is formed at the focal plane of the eyepiece by a 16-mm objective. A transparent test chart is inserted into the focal plane of the eyepiece so that the image of the sample and the test chart are viewed together. The test chart, which is shown in Fig. 11, was furnished by the research laboratories of Kodak Limited. It consists of twenty-six circular patterns, each containing ten pairs of lines having a density of 0.5. The patterns decrease in size logarithmically and are numbered consecutively.

After the apparatus was focused and the brightness was adjusted to a predetermined level with the aid of a simple photometer, the number of individual pairs of lines resolved in each pattern of the test chart was determined by the observer. The granularity of the sample was then computed by the following formula:

$$\log G_1 - \log G_2 = (N_2 - N_1)/25, \qquad (3)$$

where G_1 is the granularity of a sample obscuring half the pairs of lines for test pattern N_1, and G_2 is the granularity of the standard sample obscuring half the pairs of lines in test pattern N_2. The values of N in this equation were obtained by plotting the number of pairs of lines just resolved in each test pattern against the test-

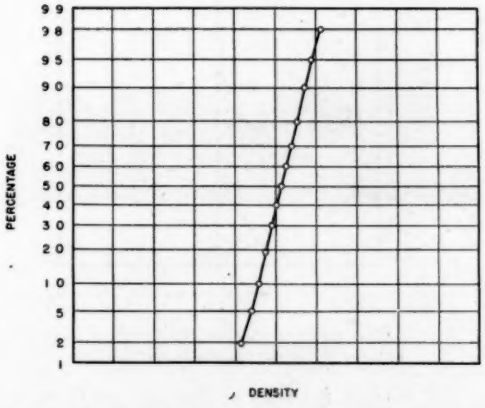

FIG. 10. Plot of an integrator record on arithmetic probability paper.

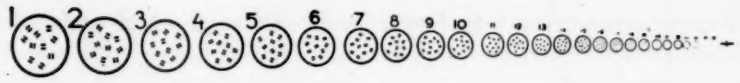

FIG. 11. Test chart used in obscuring-power apparatus to measure granularity by the Romer-Selwyn method.

pattern number, as shown in Fig. 12, drawing a curve through these points, and then noting the value of N for the condition where half the pairs of lines were resolved.

According to the theory of measuring granularity by the obscuring-power method, granularity is proportional to $1/R$, where R is the "apparent resolving power" and is measured by the dimensions of the lines in the pattern for which half the pairs of lines are resolved. The granularity of any sample can be compared to that of a standard sample by means of Eq. (3), since N is proportional to $\log R$. The ten pairs of lines in each pattern are 1.1 times the linear size of the lines in the next smaller pattern. This standard sample was furnished by Kodak Limited and had a value of 1.0 density-micron.

The values obtained by this method are shown in Table II, column H.

GRANULARITY MEASUREMENTS BY THE KÜSTER-EGGERT Q-FACTOR METHOD

The apparatus used to measure the Q factor is shown in Fig. 13. This is the instrument used by Lowry, which he described as follows:[5]

"Light from the source, X, is collimated by the lens, L_1, and passes through the aperture, A, the diaphragm, B, and an aperture, C, into the integrating sphere, S. A comparison beam is formed by the lens, L_2, and the prisms, $P_{1,2,3}$, and the magnesium oxide coated surface, D. The Martens photometer, M, is used to examine the light from the sphere and from the surface at D. In making a determination of Q, the two beams of light entering the photometer are first adjusted to equality. Then the sample is placed over the aperture, A, and the specular density, $D//$, measured. The aperture, A, is separated from C by a distance of twelve inches. This separation, together with the collimation of the incident beam and the small angle subtended at the integrating sphere by the sample, insures the specularity of the density measured when the sample is at A. The sample is then placed over the aperture, C, in the integrating sphere and the diffuse density, $D\#$ determined."

The specular and diffuse densities of both the developed photographic material and the fixed but undeveloped and unexposed material were determined for each density step on each sample. While the Q factor obtained as the ratio of the specular density to the diffuse density of the silver deposit and base is important in many cases, such as sound recording, the Q factor for the silver deposit alone is the one in which we are interested when considering the correlation between the objectively measured quantity, granularity, and the subjectively evaluated aspect of the same deposit, namely, graininess. In most cases the effect of the base may be neglected, but if the film has a gray base, the value of Q factor for the silver deposit and the base will be appreciably different from that for the silver deposit alone. For low densities, where the density of the silver deposit is of the same order of magnitude as the density of the base, to measure the Q factor without correcting for

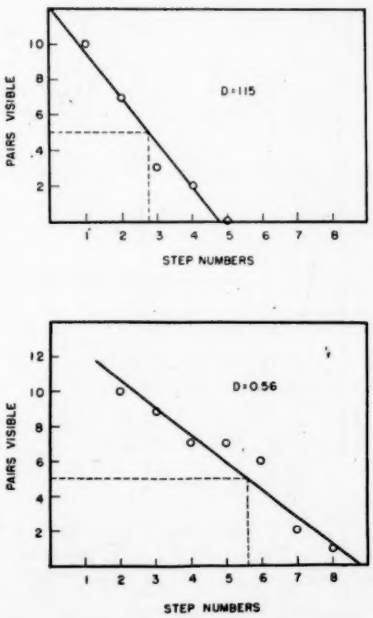

FIG. 12. Plots of the number of pairs of lines just resolved on each pattern of the test chart against the test-pattern number.

the density of the base will always lead to a value which approaches unity, irrespective of the size of the silver grains.

The Q factor was determined for each density on the several materials both with and without the correction for the density of the base. A granularity coefficient K, suggested by Küster and Eggert, was then computed by the formula

$$K = 100 \log Q,$$

where Q is the ratio of the specular density to the diffuse density for the silver deposit alone.

The results obtained by this method are shown in Table II, columns I, J, and K, columns I and J being, respectively, the values of Q factor without and with correction for the density of the base, and column K the values for 100 log Q.

DISCUSSION

Variation of Graininess with Density

Since the primary object of this investigation is to compare the results obtained by the various objective methods of measuring granularity with those given by the psychophysical method, which evaluates visual appearance, i.e., graininess, we shall first examine the graininess-density functions for all of the materials used. These functions are shown graphically in Fig. 14, in which the curves relating to the various materials are indicated by the same series of numbers as those appearing in the first column of Table I.

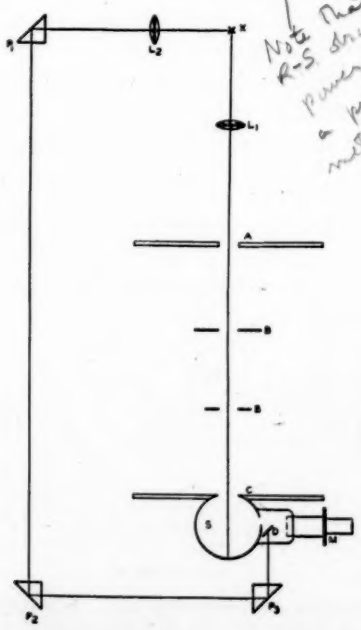

FIG. 13. Apparatus used to measure the Q factor.

It should be kept in mind that all of the densities of any one material are the result of the same development treatment. The variation in density therefore is due to differences in the exposure ($I \cdot t$) administered to the steps of the sensitometric strip. It is well known that the graininess-density curve for a series of densities obtained by variable exposure and constant development may not be the same as that for densities obtained by constant exposure and variable development.

An inspection of Fig. 14 shows that all of the materials used in this work yield graininess-density curves having the same general shape. When density is zero, graininess of course must also be zero. As density increases, graininess increases very rapidly to a maximum and then decreases slowly with further increases in density. In general, maximum graininess is reached at relatively low values of density, between 0.2 and 0.4, although with some of the materials studied the maximum is not reached until somewhat higher values of density, 0.5 to 0.6, are attained. The average density at which maximum graininess occurs, for these nine materials, is 0.36. With the equipment at present available, it is difficult to read graininess for densities much in excess of 1.3. There is little doubt, however, that as density is increased to very high values, of the order of 3.0 or 4.0, graininess must diminish still further, and as complete opacity is approached, graininess must approach zero. We can summarize the general characteristics of graininess functions by saying that they increase from zero to a maximum in the region of relatively low densities and then decrease slowly, again approaching zero, when extremely high densities are encountered.

Variation of Granularity with Density

The granularity-density relations obtained by each of the granularity methods which we have studied are shown graphically for each of the

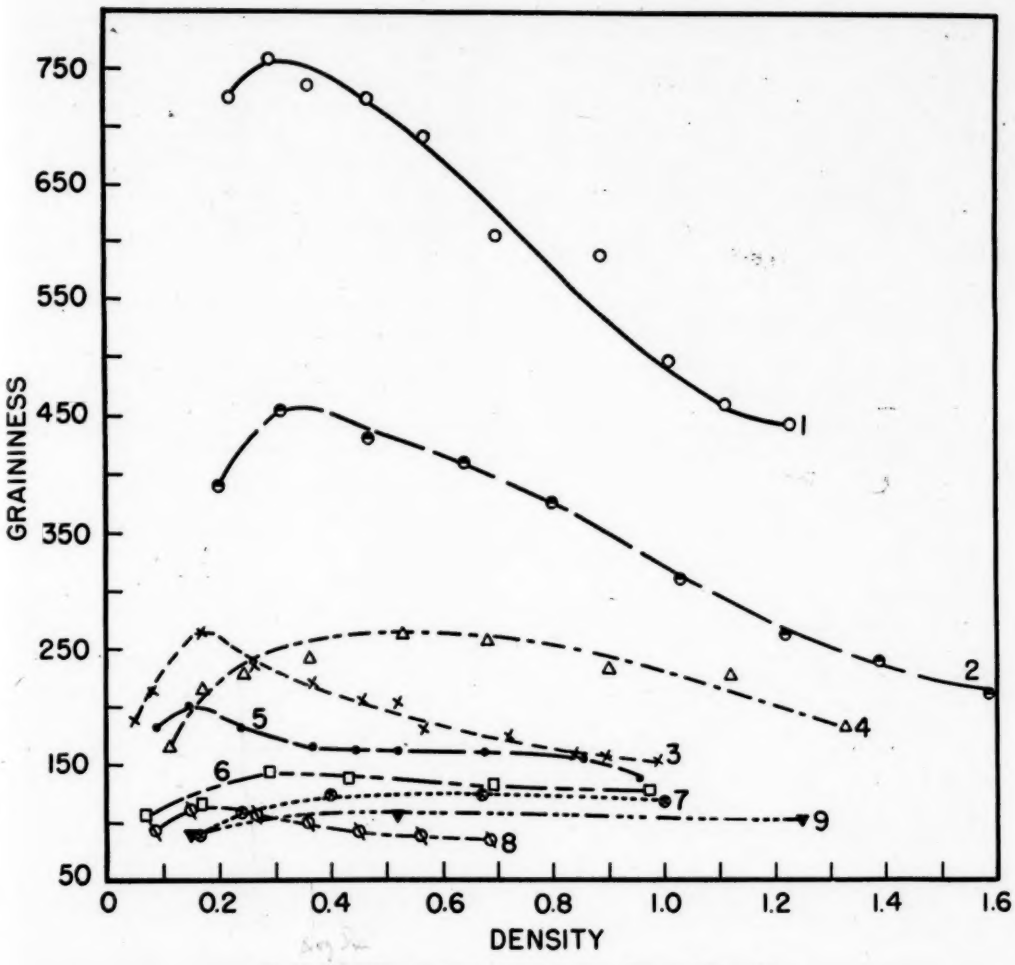

FIG. 14. Graininess-density functions for all materials used.

photographic materials in Figs. 15 to 23, inclusive. Each figure is for one of the photographic materials, as indicated by the numbers assigned to them in Table I. The graininess-density function *JD* (heavy curve) is also shown on each of these figures to facilitate the comparison of its shape with that of each of the granularity-density functions.

Since from this point on we shall have to refer frequently to the various methods of measuring granularity, it seems desirable to establish abbreviations for them. Those which have been adopted are shown in the first column of Table III. The names of the authors who are responsible for developing the various methods are given in the second column. In the last column of this table is indicated, in brief symbolic form, the basis of each method of evaluating graininess or granularity, as the case may be. The meaning of each symbol is explained at the bottom of the table.

If the graininess values for the various photographic deposits are to be directly proportional to granularity values (Graininess = Constant × Granularity), the variation of granularity with density must be identical with that for graininess. As pointed out in the preceding section, when a sample of varying density is viewed with

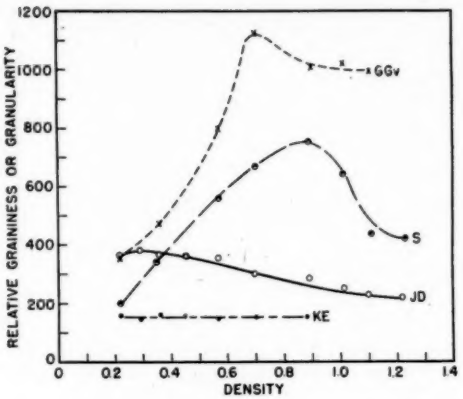

FIG. 15. Variation of relative graininess or granularity with density for material No. 1.

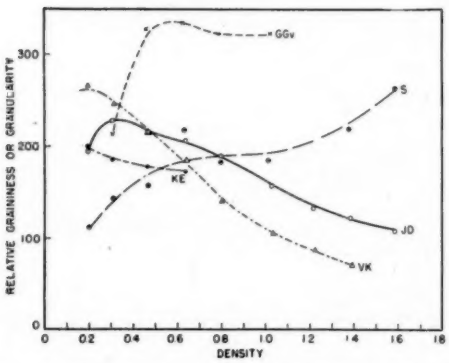

FIG. 16. Variation of relative graininess or granularity with density for material No. 2.

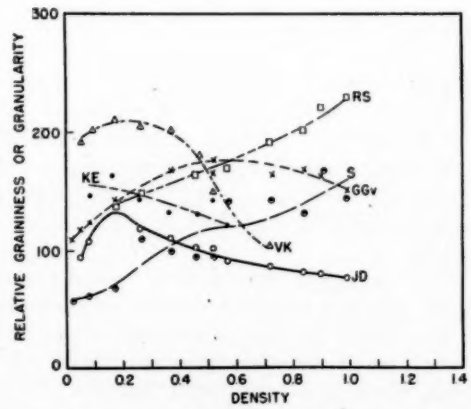

FIG. 17. Variation of relative graininess or granularity with density for material No. 3.

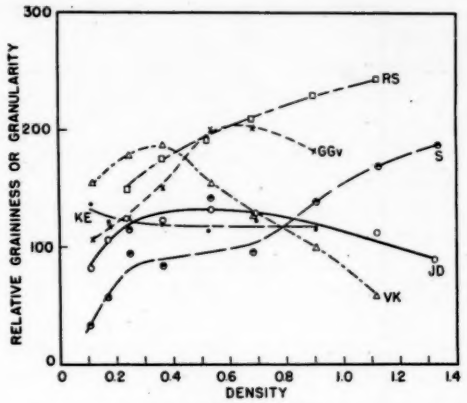

FIG. 18. Variation of relative graininess or granularity with density for material No. 4.

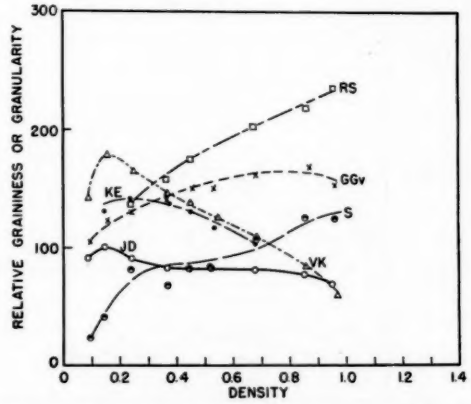

FIG. 19. Variation of relative graininess or granularity with density for material No. 5.

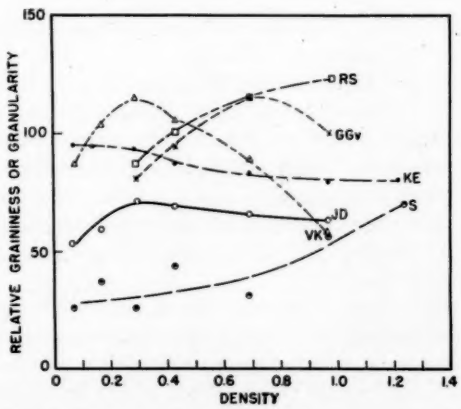

FIG. 20. Variation of relative graininess or granularity with density for material No. 6.

constant illumination, the graininess of the deposit increases rapidly with density for low densities, goes through a maximum, and then decreases for high densities, this maximum in general occurring at densities between 0.2 and 0.4.

The granularity-density curves by the Goetz-Gould method (GG_v) show some similarity to the graininess curves (JD) in that for the majority of cases they go through a maximum. However, this maximum occurs at densities between 0.5 and 0.8 rather than at densities between 0.2 and 0.4, and in the case of samples No. 7 and No. 9 (Figs. 21 and 23) no maximum value of granularity is reached in the range of densities examined. While the granularity-density curves for samples Nos. 3–6 and 8 (Figs. 17–20 and 22) go through a maximum, their general shape is markedly different from those for graininess in that as density increases granularity slowly approaches a maximum and then decreases at approximately the same rate. That is, while they are concave to the density axis, they tend to maintain a constant curvature. The granularity values for samples No. 1 and No. 2 (Figs. 15 and 16) approach a maximum rapidly but after passing through this maximum they approach a constant value which does not vary with density. The curve in Fig. 9 for GG_s is more similar to the graininess-density curves than those for GG_v, which represent granularity values which have been corrected by the visual discrimination factor to give what Goetz and Gould call "visual graininess." On the other

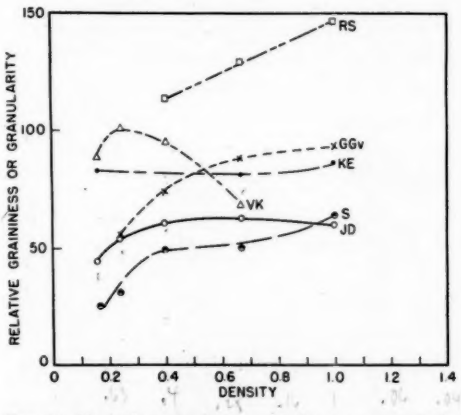

FIG. 21. Variation of relative graininess or granularity with density for material No. 7.

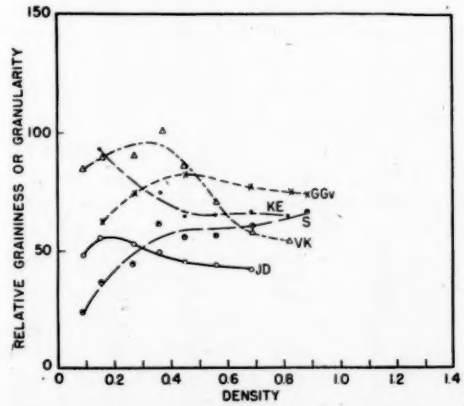

FIG. 22. Variation of relative graininess or granularity with density for material No. 8.

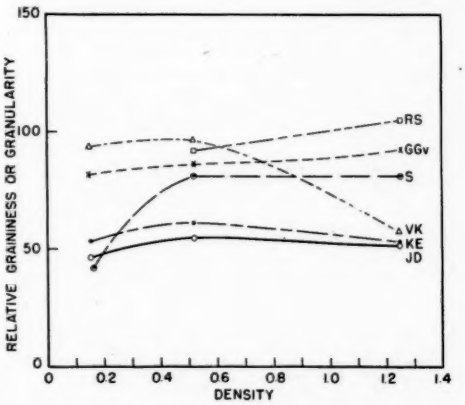

FIG. 23. Variation of relative graininess or granularity with density for material No. 9.

TABLE III. Abbreviations used to indicate the methods of measuring graininess and granularity.

Abbreviation	Authors	Basis
JD	Jones-Deisch	Blending distance
VK	Van Kreveld	$(T_1-T_2)_{Av}$
GG	Goetz-Gould	$\sigma\sqrt{2}$ for relative transmittance variations
GG_v	Goetz-Gould	$GG\times B/\Delta B$
GG_s	Goetz-Gould	$GG\times T_m$
S	Selwyn	$\sigma(2a)^{\frac{1}{2}}$ for density variations
RS	Romer-Selwyn	Obscuring power
KE	Küster-Eggert	100 log Q

σ = standard deviation
T = transmittance
T_m = mean transmittance
$B/\Delta B$ = visual brightness discrimination factor
Q = Callier factor = specular density/diffuse density

hand, Goetz and Gould state that granularity, GG_s, in terms of absolute transmittance fluctuations should be analogous to the granularity which is effective in sound recording and reproduction and not to graininess.

The granularity-density curves by the van Kreveld method (*VK*) go through a maximum at densities between 0.2 and 0.5 but the rate of decrease of granularity with density after this maximum has been reached is in every case much too great as compared to the rate at which graininess decreases. This rapid decrease in granularity with increasing density occurs, at least in part, because these measurements are in terms of absolute transmittance fluctuations. The values by the van Kreveld method therefore might be expected to correspond to the values by the Goetz-Gould method in terms of absolute transmittance fluctuations (GG_s). The fixed nature of the shape of the granularity-density functions obtained by the van Kreveld method is particularly evident from the curves for samples No. 7 and No. 9 (Figs. 21 and 23). Because of the inherent characteristics of these particular samples, the graininess values, after going through the maximum, decrease very slowly with increasing density. On the other hand, the granularity-density functions (*VK*) for these materials are not essentially different from those for the other samples and the granularity values still decrease rapidly after passing through a maximum.

The granularity-density curves for all samples examined by the Romer-Selwyn method (*RS*) show granularity increasing with density for low densities and tending to approach a limiting value as density increases. These curves do not pass through a maximum and consequently are not similar to those for graininess.

Granularity measurements by the Romer-

TABLE IV. The ratios of graininess to granularity.

No.	Density	*JD/VK*	*JD/GG*	*JD/S*	*JD/KE*	No.	Density	*JD/VK*	*JD/GG*	*JD/S*	*JD/KE*
1	0.22	—	12.50	439	22.9		.68	.992	7.59	344	10.12
	.29	—	—	—	25.1		.90	1.160	7.25	211	10.09
	.36	—	9.57	268	22.9		1.12	1.917	—	170	—
	.45	—	—	—	23.4		1.33	—	—	123	—
	.57	—	5.13	156	23.3						
	.70	—	3.08	113	20.2	5	0.09	0.657	10.22	920	—
	.84	—	3.20	105	19.5		.15	.558	9.57	670	7.73
	1.01	—	2.50	95	—		.24	.545	8.18	277	6.45
	1.11	—	2.35	132	—		.37	.572	7.22	302	6.01
	1.23	—	—	131	—		.45	.589	6.35	254	6.40
							.58	.660	6.35	254	7.17
2	.20	0.738	—	434	9.78		.68	.745	5.66	193	7.92
	.31	.929	11.67	396	12.33		.86	1.000	5.16	160	—
	.47	1.005	8.00	346	12.13		.96	1.175	4.86	141	—
	.64	1.114	7.36	235	11.98						
	.80	1.346	6.85	260	—	6	.07	.600	—	525	5.53
	1.03	1.481	5.36	214	—		.17	.548	8.85	383	5.96
	1.22	1.509	—	115	—		.29	.622	11.00	715	7.53
	1.39	—	—	138	—		.43	.662	8.69	397	8.03
	1.59	1.729	—	100	—		.69	.739	6.65	532	7.69
							.97	1.058	7.06	282	7.89
3	.05	.500	10.00	—	—						
	.08	.523	10.85	434	7.31	7	.16	.500	—	450	5.39
	.17	.633	11.57	484	8.06		.24	.550	12.22	440	5.39
	.26	.576	—	262	8.08		.40	.658	9.62	313	8.93
	.37	.551	7.69	279	8.42		.67	.893	8.33	313	7.76
	.46	.575	—	276	7.81		1.00	—	6.37	242	6.87
	.52	.736	6.87	275	6.75						
	.57	.783	5.81	157	7.47	8	.09	.565	—	480	—
	.72	.795	6.03	152	—		.15	.634	11.10	370	5.90
	.84	—	5.13	156	—		.27	.606	8.38	311	4.40
	.90	—	5.16	119	—		.36	.490	7.92	206	7.05
	.99	—	5.17	135	—		.45	.565	6.86	213	7.56
							.56	.692	5.29	200	6.43
4	.11	.535	9.76	664	6.10		.69	.672	6.00	187	6.00
	.17	.620	11.42	482	8.93						
	.24	.648	11.50	307	10.00	9	.15	.474	6.92	300	8.65
	.36	.645	9.68	372	9.84		.52	.559	7.27	168	8.79
	.53	.825	7.76	230	11.33		1.25	.883	5.58	163	9.91

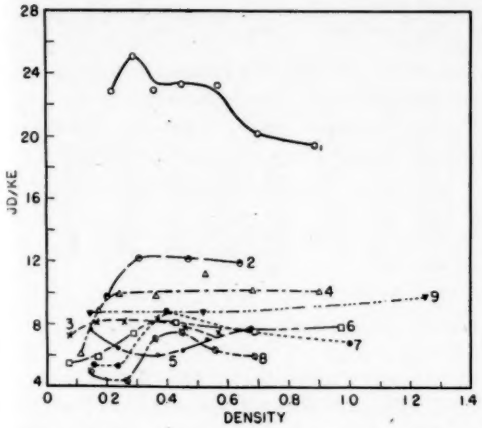

FIG. 24. Ratio of graininess (*JD*) to granularity (*KE*) as a function of density.

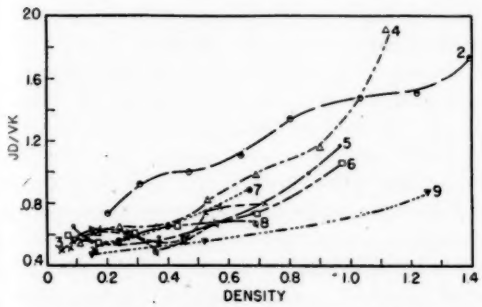

FIG. 25. Ratio of graininess (*JD*) to granularity (*VK*) as a function of density.

Selwyn method could not be made on sample No. 1 with the obscuring-power test charts available because of the very grainy structure of the silver deposit. The granularity-density curves obtained by the Selwyn method (*S*) run approximately parallel to those for the Romer-Selwyn method. However, the curve obtained by the Selwyn method on sample No. 1 (Fig. 15) is unique among the curves obtained by this method in that it goes through a maximum. With this one exception the granularity-density curves obtained by the Selwyn method do not go through a maximum and are not similar to those for graininess. Even this one curve (Fig. 15) which exhibits a maximum is not similar to those for graininess inasmuch as it slowly approaches maximum at the relatively high density of 0.9 and then decreases rapidly with increasing density.

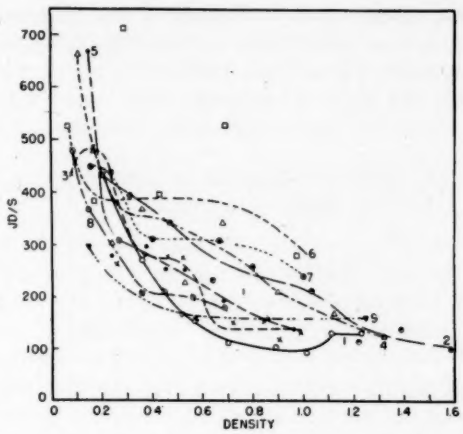

FIG. 26. Ratio of graininess (*JD*) to granularity (*S*) as a function of density.

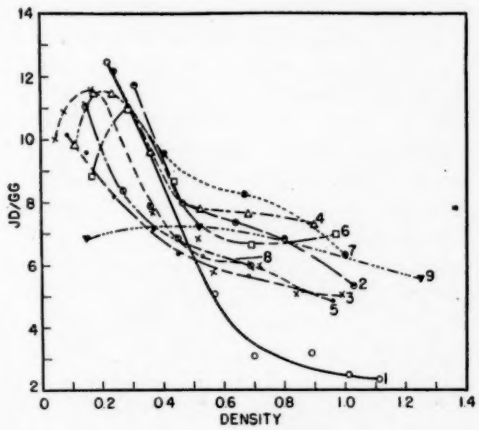

FIG. 27. Ratio of graininess (*JD*) to granularity (*GG*) as a function of density.

The granularity-density curves obtained by the Küster-Eggert method (*KE*) on samples Nos. 1, 4, 6, and 7 (Figs. 15, 18, 20, and 21) tend to be straight lines parallel to the density axis. For these samples granularity is practically independent of density for the range of densities used here. The *KE* curves for the other samples show granularity decreasing with increasing density, but, with the exception of the curve for sample No. 5 (Fig. 19), they do not go through a maximum. These curves are not similar in shape to those of graininess.

It is clear from this analysis of Figs. 15 to 23

that none of these methods of measuring granularity gives values which vary with density in the same manner as graininess and that consequently these granularity values are not directly proportional to those for graininess.

The Ratio of Graininess to Granularity as Related to Density

While none of the objective methods gives a granularity-density function in agreement with the graininess-density function, it is still possible that the ratio of graininess to granularity may be a simple function of density [Graininess/Granularity $= F(D)$]. If such a function of density exists and if it is independent of the photographic material, then the graininess of a silver deposit can be expressed as the product of granularity and the value of this function for the density corresponding to that of the deposit. This functional relationship between graininess and granularity is a psychophysical relationship and can only be established by the use of psychophysical methods. Hence if values of granularity obtained by purely objective methods are modified by the use of this psychophysical relationship, the resultant values of graininess are psychophysical quantities.

The ratios of graininess to granularity for all densities on the several samples are given in Table IV and are shown graphically in Figs. 24 through 27. Each figure shows the variation of this ratio with density for all samples as measured by a single method. The graininess-granularity ratios, given in the fourth column of Table IV and plotted in Fig. 27, are for graininess (JD) and granularity (GG) obtained by the use of the Goetz-Gould method unmodified by the visual discrimination factor. These unmodified values were used since their correction by the visual discrimination factor did not give granularity values, GG_v, which were in agreement with those of graininess. If the granularity-density function for the modified granularity values had been in agreement with those of graininess, the discrimination factor would have been the function of density which is being sought. The curves in Fig. 27 represent the exact functions which will convert the granularity values, GG, into psychophysical quantities directly proportional to graininess values. While the general shape of these curves for the various samples is similar, the differences are real. Thus, the curve for sample No. 1, for example, when used to modify the granularity values of sample No. 4 does not give psychophysical values which are directly proportional to those of graininess.

TABLE V. The ordering of the photographic materials by various methods of measuring graininess and granularity.

No.	Graininess JD Max.	JD D	KE D=0.5	VK Max.	VK D	Granularity GG_v Max.	GG_v D	S D=0.36	RS D=0.36
1	757	0.32	30.3	—	—	180	0.70	2.70	—
2	455	.33	34.6	530	0.20	53	.58	1.20	84*
3	265	.18	27.5	420	.23	28	.55	.82	60
4	265	.55	23.7	375	.36	32	.65	.70	70
5	200	.16	24.3	360	.16	26	.77	.65	63
6	143	.33	17.0	230	.29	19	.70	.25	38
7	125	.60	17.0	200	.25	11	**	.37	44*
8	111	.19	14.5	195	.33	13	.45	.45	—
9	109	.56	12	195	.52	13	**	.56	—

* By interpolation.
** Density of 0.36 (no maximum).

The curves for the variation of the graininess-granularity ratio with density shown in Fig. 26 are for the Selwyn method. While, in general, the curve shapes for the different samples are similar, the differences are real and the function of density as obtained from measurements on any one sample, when used to modify the granularity values of any other sample, does not give psychophysical values which are in agreement with those of graininess. The same conclusions apply to the curves in Fig. 25 which are derived from the granularity measurements made by the van Kreveld method and to the curves in Fig. 24 for the measurements by the Küster-Eggert method.

Special note should be taken of the curve for sample No. 1 in Fig. 24. The values of the graininess-granularity ratio for this photographic material (No. 1) are much higher than for any other of the samples measured. This indicates that graininess has increased much more rapidly than granularity (KE), as measured by the Q-factor method. While the amount of light scattered by the grains in the photographic deposit increases with grain size for most materials, there obviously must be a limit to the grain size for which this is true. Possibly the developed silver grains in sample No. 1 are so large that they have passed this limit and actually scatter less light than the smaller grains in the other samples.

Although it is apparent from Figs. 15 to 23 that graininess is not equal to the product of granularity by a constant, and from Figs. 24 to 27 that the ratio of graininess to granularity is not a unique function of density, there is evidence that some sort of general relation exists between the graininess and granularity values.

THE POSITION OF PHOTOGRAPHIC MATERIALS ON GRAININESS AND GRANULARITY SCALES

While the graininess-density curves (Fig. 14) of photographic materials are in a general way similar in shape, in that graininess increases rapidly in the region of relatively low density and then decreases rather slowly in the region of higher densities, there are, in fact, rather marked and significant shape differences. The complete graininess characteristics of a material can therefore be expressed only in the form of a curve or of a table showing the relation between graininess and density. However, for some purposes it is desirable to express the graininess characteristics of a material by a single number. If this is done and if the method of obtaining the single number to be assigned to each material is properly chosen, then these values when located on a scale of graininess will show not only the correct sequence, but also the actual relative position of any material within the series.

For this purpose it has become customary to give, as the most significant single number representing the graininess of a material, the maximum value, regardless of the density at which this maximum occurs. This practice seems to be a reasonable one since such numbers indicate the maximum extent to which negatives, consisting of an array of different densities, may be magnified without exhibiting at any point (that is, at any value of density) visually perceptible spatial inhomogeneities or granular structure. The values of maximum graininess and the densities at which they occur are shown in columns 2 and 3, respectively, of Table V.

While no method of measuring granularity gives values for the variation of granularity with density which are in agreement with graininess measurements, it may be possible to assign a single number of granularity to each of the materials (for each of the methods studied) which is proportional to the value of maximum graininess, or, at least, such that the materials will be arranged in the same order as that resulting from the use of maximum graininess.

How shall such numbers be chosen for each of the granularity methods? In one case, the Küster-Eggert method, the answer seems to be unequivocal. The proponents of this method have proposed that the single number to be used in specifying the granularity of the material be that corresponding to a constant diffuse density of 0.5. In fairness to this method, therefore, we feel that we should follow this proposal. The *KE* granularity values in compliance with the recommendation of Küster and Eggert are shown in column 4 of Table V.

The choice of the single granularity value to represent the measurements made by the van Kreveld method seems also to be directly indicated. The van Kreveld granularity-density curves rise to a maximum, then fall rather rapidly. The density at which maximum granularity occurs varies somewhat (from 0.2 to 0.5) among the materials studied. Hence the most reasonable course in the case of this method seems to be the use of the maximum. Values of maximum granularity and the corresponding densities are shown in columns 5 and 6, respectively, of Table V.

The same reasoning applies to the choice of the single value for the Goetz-Gould method (GG_v), although for two of our materials no maximum is reached within the range of densities used. Densities for maximum granularity vary from 0.5 to 0.8. The values of granularity after correction by the application of the visual brightness discrimination factor, GG_v, were used, inasmuch as the proponents of this method made this correction with the hope of obtaining numbers proportional to graininess. The maximum GG_v values and the corresponding densities are shown in columns 7 and 8 of Table V. For the two materials which do not show a maximum, the values of GG_v corresponding to 0.36 density are used.

The two remaining methods, namely, the Selwyn and the Romer-Selwyn, give granularity values which, with the exception of material No. 1 (Fig. 15), increase steadily with density. Some of these appear to be approaching a maximum at the highest densities used, but others

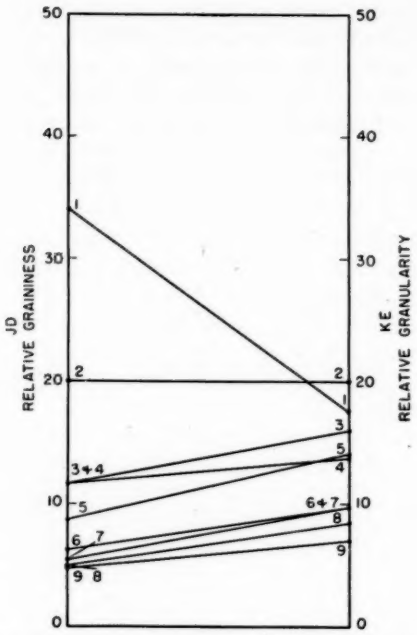

FIG. 28. Maximum graininess (*JD*) compared with a unique value of granularity (*KE*).

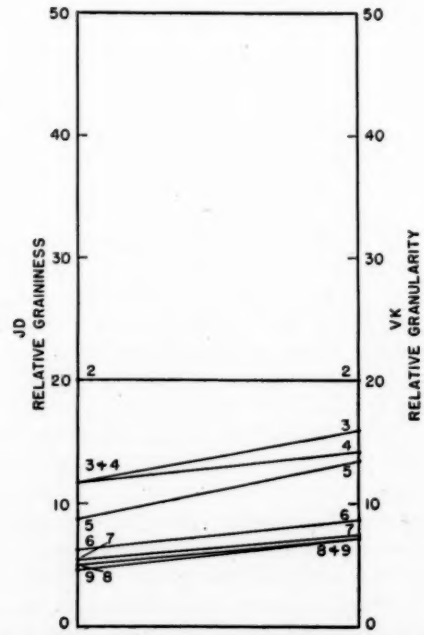

FIG. 29. Maximum graininess (*JD*) compared with a unique value of granularity (*VK*).

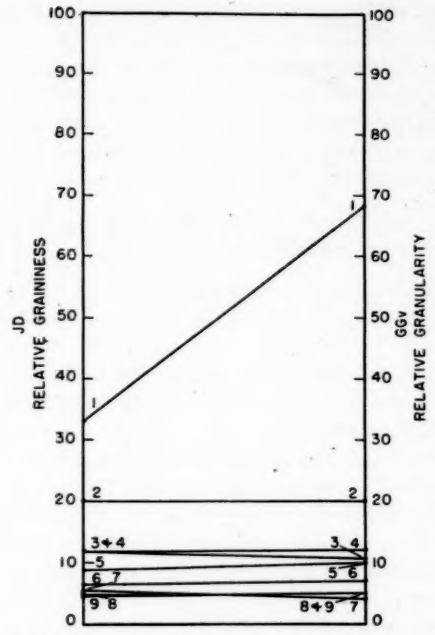

FIG. 30. Maximum graininess (*JD*) compared with a unique value of granularity (GG_v).

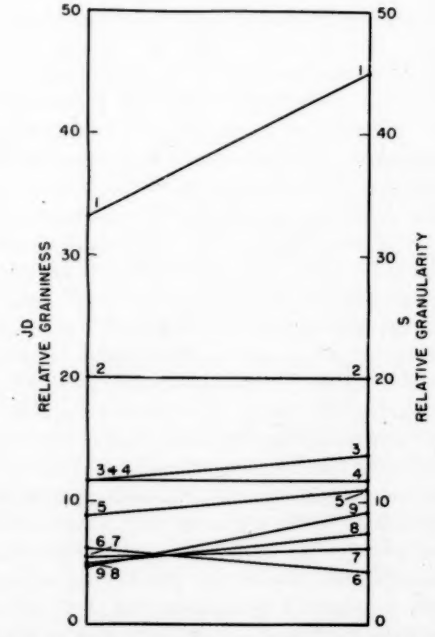

FIG. 31. Maximum graininess (*JD*) compared with a unique value of granularity (*S*).

are still increasing rapidly in the high density region. A choice of a single granularity value in these cases is therefore arbitrary. We have decided to use the granularity corresponding to a constant density of 0.36. This is the average density for the maximum graininess values of these nine materials. This seems to be reasonable since, if the granularity is to be a useful index of graininess, it should be proportional to the maximum graininess values or at least should be such as to place the material in the same sequential order as that given by maximum graininess. It may be contended that this use of a fixed density is not entirely fair to these methods and that in the case of each material, the granularity value should be taken at the density which gives maximum graininess. To do this presupposes that the density giving maximum graininess is known, but this cannot be predicted and can only be established by the determination of the graininess-density function. One of the purposes in the development of an objective method for the evaluation of the granular structure of a silver image is to avoid the necessity of making graininess measurements. The procedure we use (the use of a constant density of 0.36) is not in conflict with Selwyn's proposal regarding the assignment of a single granularity number to a negative material. He states: "It is suggested, therefore, that the graininess of a negative material be described for pictorial purposes by the granularity divided by the slope of a characteristic curve at a point 1 log exposure unit above Jones' 0.4*G* point." This proposal applies specifically to the assignment of a single number to a negative material as an indication of the graininess of prints made therefrom and is not intended as the most significant granularity value for the representation of the visual appearance of the negative as such. The granularity values for a constant density of 0.36, given by the Selwyn and the Romer-Selwyn methods, are shown in columns 9 and 10 of Table V.

An inspection of Table V shows that none of the objective methods of measuring granularity give values proportional to graininess. The magnitude of this failure can most readily be perceived by a graphic representation such as is given in Figs. 28 to 32, inclusive, in which the single or unique values for each method are shown in relative terms, these being adjusted so that all methods (both graininess and granularity) have the same numerical value for material No. 2.

In Fig. 28 it will be seen that the graininess of material No. 1 is considerably greater than that for No. 2, while the *KE* value for No. 1 is somewhat less than for No. 2. Moreover, while materials Nos. 3 and 4 have almost identical graininess, they are separated to an appreciable degree on the *KE* scale. Likewise, materials Nos. 4 and 5, which are almost identical on the *KE* scale, are separated by an appreciable distance on the graininess scale. On the whole, therefore, it must be concluded that the *KE* values are not proportional to the graininess values, and moreover actually disturb very noticeably the order in which it classifies these materials.

In Fig. 29 it will be seen that the *VK* values of materials Nos. 3 and 4 are measurably different, while they are practically identical from the standpoint of graininess. Although there are no actual reversals of position on the *VK* scale, the

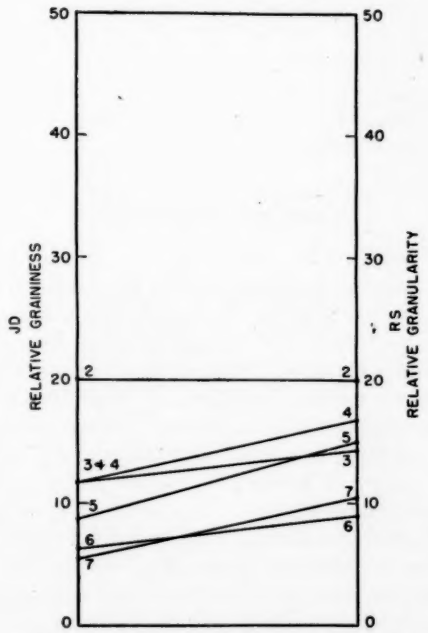

FIG. 32. Maximum graininess (*JD*) compared with a unique value of granularity (*RS*).

locations are rather badly distorted compared with the graininess scale.

In Fig. 30 it will be seen that the granularity value, GG_v, for material No. 1 is excessively high compared with its graininess value and that materials Nos. 3 and 4, which have the same graininess, are separated by a measurable difference in granularity. The remainder of the materials show relatively little displacement with respect either to order or to position.

In Fig. 31 the *S* value for material No. 1 falls unduly high on the granularity scale compared with the graininess scale. Materials Nos. 3 and 4, which are identical in graininess, are displaced by a measurable amount in terms of granularity, and material No. 9, which has the lowest graininess value, has been moved up in granularity to a position above materials Nos. 6–8. Thus, the proportionality is not very good and for samples of low graininess the disturbance of order is rather marked.

In Fig. 32 the *RS* values for materials Nos. 3 and 4 are again separated appreciably on the granularity scale, although they are identical with respect to graininess. Material No. 5, whose graininess is appreciably less than that of No. 3, is placed by the *RS* granularity value slightly above material No. 3. Thus, the proportionality is not very good and there is a marked disturbance of the order in which they are placed on the graininess scale.

A COMPARISON OF THE METHODS OF MEASURING GRANULARITY

The *GG*, *VK*, and *S* values of granularity are obtained by techniques which are similar in that all are based on fluctuations in the amounts of light transmitted by small areas of the silver deposit. In view of this, it seems reasonable to expect a rather high degree of linear correlation between the values obtained by these methods. However, it should be noted that the *VK* values are given in terms of absolute transmittance fluctuations; the *GG* values, in terms of relative transmittance fluctuations; and the *S* values, in terms of density fluctuations. Since relative transmittance fluctuations ($\Delta T/T_m$) are an approximation to the density fluctuations, ΔD, the *GG* and *S* methods can be compared directly. By using the *GG* and *S* values of granularity for all the densities on all the materials used in this work, the coefficient of correlation between the *GG* and *S* values was found to be 0.96. To make the *VK* values of granularity in terms of absolute transmittance fluctuations comparable with the *S* values of granularity in terms of density fluctuations, all *VK* values were divided by the mean transmittance of the deposit. The coefficient of correlation between these modified *VK* values and the *S* values was found to be 0.84.

The proponents of the *RS* and of the *S* methods report that these techniques yield values which are closely related to each other. The coefficient of correlation between the values which we have obtained by these methods was found to be 0.86.

While the correlation coefficient between the *GG* and the *S* values is 0.96, it should be recognized that the linear relation between these two variables is of little use for converting the granularity values obtained on a specific photographic material by one method into values that are comparable to those obtained by the other. In Fig. 33 the *GG* and *S* values have been plotted against each other for all densities on five of the materials used (Nos. 2, 3, 5, 6, and 8). The linear relation between these two sets of values is shown in this figure by a straight line which was obtained by the method of least squares. This line was calculated by using the granularity values obtained on all nine samples. Two curves, Nos. 2 and 3, have been drawn through the measured points for samples Nos. 2 and 3. These curves depart widely from the straight line and there is little doubt that, for these particular samples, they represent more precisely the relation between the granularity values made by the two methods under consideration than does the straight line.

CONCLUSIONS

When a sample of varying density is viewed with constant illumination, the graininess of the deposit increases with density for low densities, goes through a maximum, and then decreases for high densities. Since none of the granularity-density functions is in agreement with the graininess-density functions, graininess is not a linear function of granularity as measured by any of the methods studied.

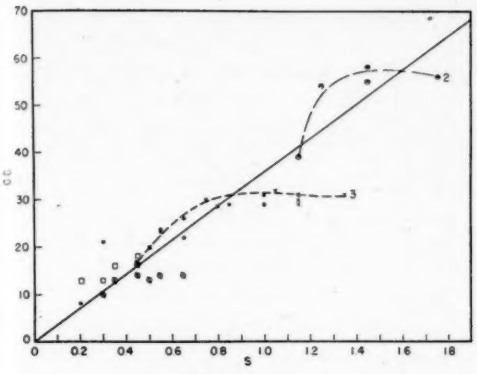

FIG. 33. Granularity (*GG*) plotted against granularity (*S*), using all densities on materials Nos. 2, 3, 5, 6, and 8.

Inasmuch as the ratio of graininess to granularity is not the same function of density for different materials, graininess in general cannot be computed as the product of granularity and a function of density which is independent of the material.

All methods of measuring granularity which were investigated rank photographic materials in the same general order as graininess measurements, but for no method is the order exactly the same and, in some cases, the differences in order are very marked. The measurements, while in general qualitative agreement as to order, are not in good quantitative agreement with graininess measurements. The fact that the various methods of measuring granularity rank the materials roughly in the order of graininess indicates that they measure certain factors controlling graininess; on the other hand, the fact that these methods fail to yield granularity values which vary with density in the same manner as graininess values indicates that they do not measure all the factors controlling graininess.

The van Kreveld, Goetz-Gould, and Selwyn methods are similar in that granularity is evaluated in terms of fluctuations in the amount of light transmitted by small elements of area. However, measurements by one method cannot be used to predict the values which would be obtained by the other methods.

Uses of Retardation Plates in Spectrophotometry

I. Determination of Slit Shape and Slit Width

GEORGE L. BUC AND E. I. STEARNS
Research Department, Calco Chemical Division, American Cyanamid Company, Bound Brook, New Jersey

(Received April 6, 1945)

A retardation plate is used to make a "calibrator" having a known transmission curve. It is shown how the effect of slit width and slit shape on the plotted transmission of the calibrator can be predicted and conversely how the width and shape of an assumed slit may be evaluated from the performance of a spectrophotometer in plotting the transmission of the calibrator. If it is assumed that the slit is trapezoidal, it is recommended that it be described on an over-all performance basis in terms of two factors, H the slit width and f the shape factor. Regardless of the assumed slit geometry it is probable that two spectrophotometers that give similar transmission data for a calibrator will give similar curves for other types of samples.

As greater accuracy is required of spectrophotometric data, it becomes more necessary to define conditions of measurement. It is, therefore, becoming of increasing interest to be able to evaluate slit width.[1] In recent government specifications involving analysis of colored material with spectrophotometers, it is specified that measurements should be carried out with a ten-millimicron slit, or, if a thirty-millimicron slit is used, a correction should be applied to the data. This is an indication of a general trend toward the recognition of the fact that if data from different laboratories and from different spectrophotometers are going to be compared, it is necessary to specify the exact conditions of measurement, and this includes a description of slit used.

A method was developed in this laboratory for rapidly evaluating slit width on a performance basis with the assumption of a symmetrical "triangular" slit. It then became evident that the assumption of a symmetrical "triangular" slit was a rather questionable one, and a somewhat longer method was developed for evaluating the slit based on an assumed symmetrical "trapezoid" shape. This resulted in a more precise slit description and permitted spectrophotometer behavior to be more exactly predicted in many cases. The slit was then described with a width and a shape factor. If some behavior were still not satisfactorily predicted, it would be possible to assume asymmetric slit shapes and develop such cases along similar lines. Even without specific assumptions as to slit geometry, this method gives an empirical measure of slit geometry which is useful in comparing the performance of different instruments.

There has been a general tendency to simplify discussions of slit, and, rather than presenting the entire intensity distribution curve, to describe the slit in terms of "slit width." Recent discussions of slits have been based usually on the presentation of Hogness.[2] Slit width as used here is defined as the difference in wave-lengths between the half-maximum intensity points on the slit intensity distribution curve. This is one half of the "spectral region isolated" as defined by Hogness. This simplification introduces an uncertainty because it does not completely define the slit as is shown in the following discussion.

Figure 1 illustrates a theoretical slit-width effect calculated by methods which are discussed later. If a particular sample were measured twice, the only difference in the two measurements being slit width, two widely different curves could be obtained.

Figure 2 illustrates a theoretical slit-shape effect with symmetrical slits also calculated by methods discussed later. If a particular sample were measured twice, the difference in the two measurements being only slit shape, a variable not often defined, two rather widely different curves could be obtained.

There is a tendency to define slit width only

[1] M. G. Mellon, Ind. Eng. Chem. Anal. Ed. **17**, 81–88 (1945).

[2] T. R. Hogness, Z. P. Zscheile and A. E. Sidwell, J. Phys. Chem. **41**, 379–415 (1937).

in terms of the intensity distribution of the light leaving the monochromator, although in experimental work it is necessary to use a performance definition.[3] Actually, if the photometer alters the effective intensity distribution, for instance by unequal response with wave-length of the photo-cell, the effect on the plotted absorption curve appears to be a slit-width effect. Thus a filter photometer using a tungsten light source and a Wratten 89A filter may be said to have a "slit width" of several hundred mu. But if the photometer includes an RCA 929 photo-cell, all energy of longer wave-length than 750 mu is ineffective, and the "effective slit width" is less than 100 mu. Moreover, the center of gravity of the effective slit will be further in the visual than expected if the photo-cell is not considered. It is, therefore, desirable to measure slit widths on a performance basis and include the effect of factors outside the monochromator. Figure 3 presents the slit shape assumed in the following discussion. The ordinate is response rather than intensity in order to describe the slit on a performance basis.

A transmission sample having an absorption curve in which the transmission as a function of wave-length is known theoretically and independently of any spectrophotometer determination may be used to measure slit shape and slit width on a performance basis. Polarizing optics are used to obtain such a curve. One optical combination used consists of two crossed Nicols with a retardation plate inserted between them at an angle of $\pi/4$ radians to the axes of the Nicols. This or similar combinations will be called a "calibrator" and is shown in Fig. 4.

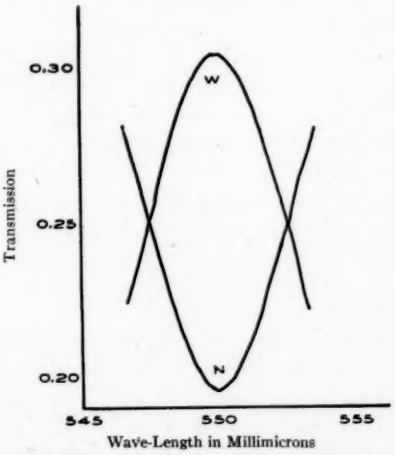

FIG. 1. Illustration of theoretical slit-width effect: calibrator with 3.30-mm thick quartz plate; N is with slit width 8.12 mμ, shape factor 1; W is with slit width 14.25 mμ, shape factor 1.

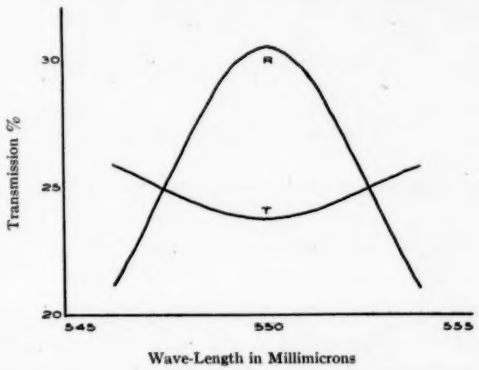

FIG. 2. Illustration of theoretical slit-shape effect: calibrator with 3.30-mm quartz plate; T is with slit width 14.25 mμ, shape factor 0; R is with slit width 14.25 mμ, shape factor 1.

Basically the same polarizing optics can be used for slit calibration, wave-length calibration, and measurement or elimination of dichroism. The optics will be treated briefly, first, and then considered in more detail in relation to the separate functions they may perform.

Consider three optical elements traversed by parallel light in the following sequence, Nicol prism-retardation plate-Nicol prism, as shown in Fig. 4. If losses owing to absorption, scattering, incomplete polarization, and surface reflection are ignored the intensity of monochromatic light transmitted by the combination is given[4] by Eq. (1).

$$I = I_0[\cos^2(A-B) - \sin 2A \sin 2B \sin^2(D/2)], \quad (1)$$

where I_0 = intensity of light leaving N_1 (see Fig. 4); I = intensity of light leaving N_2 (see Fig. 4); A = angle from the optic axis of the retardation plate to the polarizer; B = angle from the optic axis of the retardation plate to the analyzer; and

[3] R. A. Oetjen, Chao-Lan Kao, H. M. Randall, Rev. Sci. Inst. **13**, 515–23 (1942).

[4] M. E. Mascart, *Traite d'Optique* (Gauthier-Villars, Paris, 1889), **II**, p. 22.

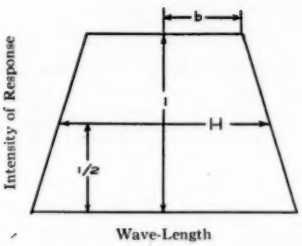

FIG. 3. Assumed distribution of intensity of response among the wave-lengths in a slit.

the retardation angle D is given by Eq. (2),

$$D=2\pi t\frac{(n_e-n_o)}{\lambda}\text{ radians,} \qquad (2)$$

where t=thickness of the retardation plate; n_e and n_o=extraordinary and ordinary indices of refraction (assumed constant over a slit-width interval); and λ=wave-length of monochromatic light in the same units as t. In the case of a calibrator Eq. (1) reduces to Eq. (3) If we let $I/I_m=T_c$, where T_c is transmission of the calibrator and I_m is intensity of collimated unpolarized light incident on N_1

$$T_c=\tfrac{1}{4}-\tfrac{1}{4}\cos D. \qquad (3)$$

It is evident from Eq. (3) that the transmission curve has a series of absorption maximums throughout the spectrum at intervals of 2π radians.

In practical spectrophotometry the plotted curve deviates from that predicted by Eq. (3) for seven reasons: absorption, scattering, surface reflection, instrument polarization, incomplete calibrator polarization, stray light effects, and finite slit width.

The combined effects of absorption, scattering, and surface reflectance losses is the same as the effect of placing an absorbing filter ahead of a calibrator free from these defects. To correct the plotted curve for these effects, the exact value of the equivalent absorption effect need not be determined since the ratio of plotted transmission of any point to double the average plotted transmission gives the true transmission at the point, the correction factors cancelling out in the ratio.

The effect of instrument polarization is to multiply the expected transmission by a factor which may vary from 0 to 2 depending on the degree of polarization of the light beam of the spectrophotometer and its orientation relative to the calibrator. The correction for this effect is applied in the process of correcting for the absorption factors described above.

The effect of incomplete polarization is to divide the light into two parts of which one is transmitted unaffected. The other part is governed by Eq. (3). To correct the plotted curve for this effect is not simple, and makes the use of imperfect polarizers inconvenient for the calculation of absolute data. Such imperfect polarizers may, nevertheless, be useful for empirical comparisons of performance. To determine the completeness of polarization, the transmission of the crossed polarizing members of the calibrator may be measured and divided by the transmissions of the parallel polarizing members. Nicols are recommended for absolute data.

The effect of finite slit width on the plotted transmission of the calibrator is discussed below. In this discussion it is assumed that complete polarizers are being used and that corrections for the four absorption effects have been made. Stray light is considered as an integral part of the response distribution curve as a function of wave-length. To the extent then that stray light affects the plotted transmission of the calibrator it is treated as a slit characteristic.

Deviations from Eq. (3) occur because the light I_m is not monochromatic but is distributed over a range of wave-lengths. For the purpose of immediate discussion the shape of the intensity of response distribution curve is arbitrarily assumed to be a trapezoid as shown in Fig. 3. H is the definition of slit width. If $b=H/2$ the slit shape of Fig. 3 is rectangular, if $b=0$, triangular, and for intermediate values of b trapezoidal.

Wave-length may be replaced by radians in Eq. (3) using the value of D corresponding to the

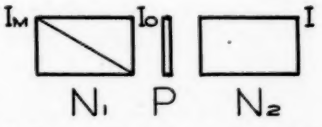

FIG. 4. Calibrator: N_1 and N_2 are crossed Nicols; P is retardation plate at $\pi/4$ radians to axes of Nicols; I_m=intensity of light from monochromator; I_0=intensity of light leaving N_1; I=intensity of light leaving N_2.

given wave-length calculated by means of Eq. (2). Assuming that D is linear in λ over the slit interval b and H may also be expressed in radians, as is D, rather than in wave-length units. For instance if the slit width extends from 500 to 510 mu the values 500 and 510 are substituted in Eq. (2) and the values of D corresponding to each is found. The difference between the two values of D so calculated is the width of the slit in radians. This means that the numerical measure of slit width, expressed in radians, depends on the thickness of the retardation plate used, and that the abscissa of Fig. 3 may be changed from λ to D. Thus the plotted transmission (T) is,

$$T=\int \frac{I_r T_e dD}{I_r dD}. \tag{4}$$

The effective intensity of response of a spectrophotometer at a particular wave-length setting in the absence of a sample is,

$$\int I_r dD=\int_{D'-b}^{D'+b} dD+\int_{D'+b}^{D'-b+H} \frac{D-D'+b-H}{2b-H}dD + \int_{D'+b-H}^{D'-b} \frac{D-D'-b+H}{H-2b}dD, \tag{5}$$

where I_r is the intensity of response as plotted in Fig. 3, D' is the retardation angle at the wave-length of interest, and H and b (expressed in radians) are defined by Fig. 3.

$\int I_r dD$ is the area of Fig. 3.

Combining Eqs. (3) and (5) as indicated by Eq. (4) the general equation relating the variables is found to be

$$T=\frac{1}{4}+\frac{1}{2}\cos D\frac{\cos(b-H)-\cos b}{H(H-2b)}. \tag{6}$$

From Eq. (6) it is evident that transmission depends on slit shape. Throughout the calculations the term b is used as defined by Fig. 3. For purposes of discussion, a slit shape factor f is used where f is defined by the equation

$$f=2b/H.$$

Two cases of interest are the transmissions at the points $D=2n\pi$ radians (n integral) with rectangular slit ($b=H/2$) and a triangular slit ($b=0$). Eq. (6) reduces in these cases to the rectangular case,

$$T_{0r}=\frac{1}{4}+\frac{1}{2}\frac{\sin H/2}{H}, \tag{7}$$

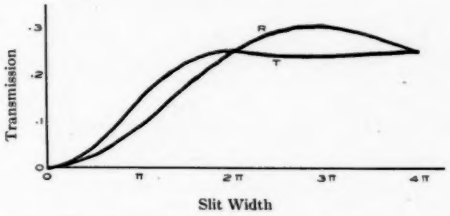

FIG. 5. Slit-characterization curves: T is curve for shape factor of 0; R is curve for shape factor of 1; 2π corresponds to the distance between two absorption peaks.

and the triangular case

$$T_{0t}=\frac{1}{4}+\frac{\sin^2 H/2}{H^2}, \tag{8}$$

where T_0=transmission of calibrator with $D=2n\pi$ radians. Eqs. (7) and (8) are plotted in Fig. 5.

It is evident from the above that the same transmission sample (a calibrator) with the same slit width (2.86π radians) will plot differently depending on the slit shape. This is illustrated in Fig. 2. The transmission data of Fig. 2 are plotted on the basis of Eq. (6) with $b=H/2$ (curve R), $b=0$ (curve T) and with $H=2.86\pi$ radians in both cases. It is also evident from Eq. (6) that transmission varies with slit width. The transmission data of Fig. 1 are plotted on the basis of $H=2.86\pi$ radians (curve W) and 1.624π radians (curve N) with $b=H/2$ in both cases.

An absorption maximum at one slit width may appear as a transmission maximum at another slit width. (See Fig. 1.) This fact was apparently neglected by Runge,[5] Paschen,[6] Weniger,[7] Muller,[8] and Hyde.[9] Their reasoning

[5] C. Runge, Zeits. Math. **42**, 205–213 (1897).
[6] F. Paschen, Wied. Ann. **60**, 712 (1897).
[7] W. Weniger, *Measurement of Radiant Energy* (McGraw-Hill Book Company, Inc., New York, 1937), p. 182.
[8] C. Muller, U.S. Patent 1,943,898 (1934).
[9] E. P. Hyde, Astrophys. J. **35**, 237–267 (1912).

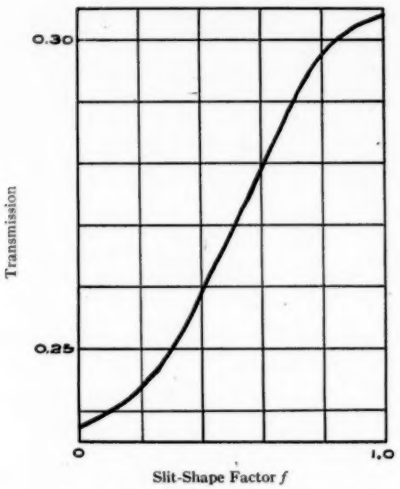

FIG. 6. Slit-shape evaluation curve.

holds only if the slit width is narrower than the absorption band to begin with.

If it is believed that slit width is enough of a description of the monochromating performance of a spectrophotometer, and it is believed that the wave-length calibration of the spectrophotometer is correct, it is possible to measure two absorption maximums of only one calibrator, assume a slit shape factor of $f=0$, using curve T of Fig. 5 read off the slit width in radians corresponding to the transmission of one of the observed absorption maximums, and convert radians to millimicrons with the information that two absorption maximums of the calibrator (which may be read from the instrument scale in millimicrons) differ by 2π radians.

Since the plot of a calibrator will depend on both slit width and slit shape, a single measurement of one calibrator cannot yield enough information to define both these factors. If, for samples with sharp peak absorptions, it is wished to describe the slit somewhat more accurately and avoid ambiguities which may be as large as those of Fig. 2, it is possible to work with several quartz plates. For instance, assume the simple slit shape of Fig. 3 and work with three quartz plates to get seven points on a slit shape characterization curve as described below. If it is believed that the slit differs from the simple shape of Fig. 3, it is possible to assume some other distribution and this assumption, together with the known transmissions of the calibrator, can be used to get a performance measurement. For instance, an asymmetric slit would fall in this class. Irrespective of the slit shape assumed, the performance of the spectrophotometer with a calibrator is an empirical measure of slit characteristics.

If it is desired to measure the slit of a spectrophotometer at a given wave-length assuming the distribution of Fig. 3, a series of quartz plates should be provided such that a range of slit widths up to 3π radians can be covered. A convenient series of plates is three in the approximate ratio of thickness of 1, 2, and 4, the greatest thickness being $\lambda^2/0.009H'$, where H' is the slit width in the same units as λ. For the value of H', the manufacturer's statement or a calculation based on the dispersion and the physical slit width may be used as a preliminary estimate. The plates are then used separately and in combination as the retardation plate of the calibrator.

The transmission curve, Eq. (3), for each calibrator has absorption maximums distributed throughout the spectrum. Each calibrator may have an absorption maximum close to any desired wave-length if the thickness is chosen as described above. The absorption maximum of the various calibrators do not in general coincide. Therefore, in measuring the slit at a given wave-length, it is necessary to measure the transmission for each calibrator at the absorption maximum of that calibrator closest to the wave-length of measurement. It is to be noted that this absorption maximum may be plotted as a transmission maximum. The transmission becomes 0.25 (average transmission for ideal case) when the slit is 2π radians wide for a symmetrical slit regardless of its shape as is found by substituting $H=2\pi$ radians in Eq. (6).

The transmission of the absorption peak for each calibrator is then plotted against the plate thickness. Using the curves of Fig. 5 as a guiding envelope pattern, a curve is drawn through the plotted points and the thickness corresponding to 25 percent transmission ($t_{0.25}$) is determined by interpolation. This value, along with the wave-length and refractive index data is sub-

stituted in the following equation to convert slit width back to wave-length units.

$$H' = \lambda^2 / t_{0.25}(n_e - n_o). \quad (9)$$

This equation is derived from Eq. (2) as outlined above and becomes, for quartz in the visual range, and t, H', and λ all expressed in millimicrons,

$$H' = \lambda^2 / 0.009 t_{0.25}. \quad (10)$$

The height of the curve in the region between slit widths of 2π and 4π gives a good indication of the slit shape since rectangular slits give a value of T_0 as high as 0.3 in this region while with triangular slits a value of $T_0 = 0.25$ is never exceeded. The relation between transmission and slit shape factor at the slit width of maximum discrimination may be deduced from Eq. (6) by letting $D = 0$ and $H = 2.86\pi$ whence

$$T_b = \frac{1}{4} + \frac{1}{2} \frac{\cos(b - 2.86\pi) - \cos b}{2.86\pi(2.86\pi - 2b)}, \quad (11)$$

where T_b = transmission at $H = 2.86\pi$. For convenience this relation has been plotted in Fig. 6.

It may be noted that this method of calibration of slit is independent of wave-length calibration.

Asymmetric slits which are outside the scope of this paper may be recognized by the fact that the transmissions in the slit-shape characterization curve of Fig. 5 do not reach 0.25 for slits less than 3π. Asymmetric slits may be treated by an extension of the method used for symmetrical slits.

One reason why the slit width may be increased over that calculated from the dispersion and the physical slit width is that the physical curvature of the exit slit may not coincide with the curvature of the image of the entrance slit. This curvature may be checked in a spectrophotometer employing a continuous source by placing the calibrator in front of the exit slit and observing the dark and bright isochromatic lines. The curvature of these lines and the physical slit should agree. The customary calculations[10] can thus be avoided.

[10] B. T. Barnes, *Measurement of Radiant Energy* (McGraw-Hill Book Company, Inc., New York, 1937), p. 176.

When it appears that width is inadequate to describe a slit, it is recommended that slits be described by two factors, the width factor H, similar to that now in use but on a performance basis, and a new factor f, the shape factor. As may be seen from Fig. 3, H is the half-height width and f is the ratio of the flat region of maximum response to H. These may be evaluated as described above by a plot of T_0 *versus* t and the help of Eq. (9) and Fig. 6.

For the special case of the Martens photometer, employed in many polarized beam spectrophotometers, a modified calibrator is recommended. The retardation plate is placed between the photometering Rochon prism and the Wollaston prism with the axis at $\pi/8$ radians to the axis of the Wollaston. This is mechanically preferable to the equivalent position of $\pi/8$ from the Rochon axis.

For the above case, but with the angle $\pi/8$ not specified, the monochromatic light in the sample and reference beams is respectively

$$I_s = \cos^2(A - B) - \sin 2A \sin 2B \sin^2 D/2, \quad (12)$$

$$I_c = \sin^2(A - B) + \sin 2A \sin 2B \sin^2 D/2, \quad (13)$$

as may be derived from Eq. (1) when the photometering Rochon and the Wollaston are considered as the polarizing and analyzing elements respectively of the calibrator.

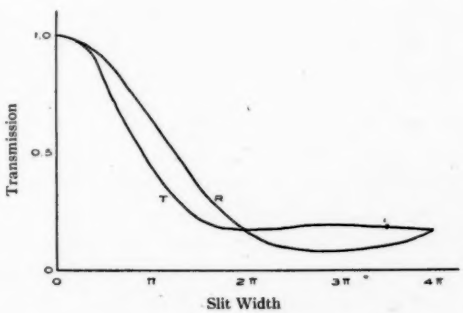

FIG. 7. Slit-characterization curve—special case of Martens photometer: T is curve for shape factor of 0; R is curve for shape factor of 1.

When the two beams are equal, the plotted curve T can be derived from Eqs. (5), (12), and

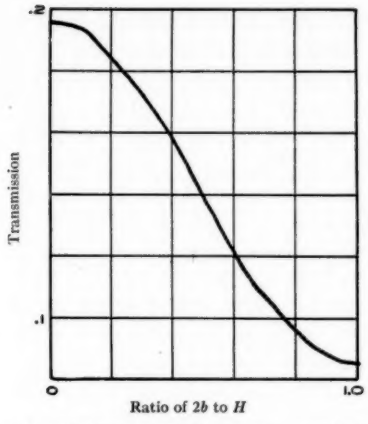

FIG. 8. Slit-shape evaluation curve—special case of Martens photometer.

(13) and is found to be

$$T=\tan^2\left[-B+\frac{1}{2}\tan^{-1}\left(\frac{H(H-2b)\cot 2B}{2\cos D[\cos(b-H)-\cos b]}\right)\right]. \quad (14)$$

In the special cases of rectangular and triangular slits, and $B=\pi/8$ (which makes T remain positive), Eq. (14) reduces to

$$T_{ir}=\tan^2\left[-\frac{\pi}{8}+\frac{1}{2}\tan^{-1}\left(\frac{-H/2}{\sin H/2}\right)\right], \quad (15)$$

$$T_{it}=\tan^2\left[-\frac{\pi}{8}+\frac{1}{2}\tan^{-1}\left(\frac{-H/2}{\sin H/2}\right)^2\right], \quad (16)$$

where T_{ir} is the transmission of the transmission maximum for a rectangular slit and T_{it} is the same quantity for a triangular slit. Equations (15) and (16) are plotted in Fig. 7 and are the slit-shape characterization curves for the calibrator in this position. The slit width is determined by substituting the thickness t corresponding to 0.171 transmission, in Eq. (17)

$$H'=\lambda^2/t_{0.171}(n_e-n_o). \quad (17)$$

This value, 0.171, was found by letting $H=2\pi$ radians in Eq. (14). The slit-shape factor is found with the aid of Fig. 8.

The accuracy in determining slit width depends on how many points are obtained in the curve which is compared with the slit-shape characterization curves.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 35, NUMBER 7 JULY, 1945

Uses of Retardation Plates in Spectrophotometry

II. Calibration of Wave-Length

GEORGE L. BUC AND E. I. STEARNS
Research Department, Calco Chemical Division, American Cyanamid Company, Bound Brook, New Jersey

(Received April 6, 1945)

It is shown that, based on the criterion that a correctly calibrated spectrophotometer will give the same transmission curve for straight-line absorptions as would be obtained with monochromatic light, the calibrator of Part I may be used for wave-length calibration with an error of less than 1 percent of slit width. Specific recommended calibration procedures are given.

THE most widely used methods for initial calibration of prism spectrophotometers in the visual and near infra-red comprise measurement of the absorption curve of a didymium filter and determination of the location of line spectra.[1–5] The use of mercury lines in conjunction with filters has also been proposed.[6] For daily calibration the didymium filter alone is used.[3]

The present paper proposes a combination of polarizing optics to calibrate wave-length. Wave-length calibration by polarization was proposed by Mouton[7] and Carvallo[8] in 1879 and in 1893, but the method has been generally overlooked except for photoelastic experiments,[9] and its advantages in modern spectrophotometry have not been previously pointed out.

A true absorption curve is a plot which would be obtained if monochromatic light were used. In deciding what is the best calibration procedure it might be thought at first that the calibration should result in the instrument plotting true absorption curves. However, it is impossible for a spectrophotometer operating with a continuous source to be so calibrated that it will yield true absorption curves for most samples. For instance an absorption maximum will be plotted with too high a transmission as a result of finite slit width. The nearest approach to a true curve requires in general a different wave-length calibration for each sample. This is because in general two crossing curves do not have the same integrated transmission (see Eq. (4), Part I) over an interval about the crossing point. It thus becomes necessary to calibrate the machine for a particular type of sample. Van den Akker[6] has shown that the same calibration is required for all straight-line absorption curves, provided that the straight line extends over a wave-length interval wider than the slit width. Calibration by the Van den Akker method results in true absorption curves being plotted for the straight line intervals. The absorption curves of many samples differ but slightly from straight lines, and thus the true curve is closely approximated.

In view of the above, the correct wave-length calibration shall be defined as that which results in true absorption curves being plotted for long straight-line intervals of absorption curves.

Basically the same polarizing optics which were proposed for slit calibration, in Part I of this series and which were called a "calibrator" for brevity can be used for wave-length calibration. In the absence of absorption, scattering, surface reflection, instrument polarization, and incomplete calibrator polarization, as was shown in Eq. (6), Part I, the plotted transmission (T) of the calibrator is

$$T=\frac{1}{4}+\frac{1}{2}\cos D\frac{\cos(b-H)-\cos b}{H(H-2b)}, \qquad (1)$$

[1] W. R. Brode and C. H. Jones, J. Opt. Soc. Am. **31**, 748 (1941).
[2] H. H. Cary and A. C. Beckman, J. Opt. Soc. Am. **31**, 682–89 (1941).
[3] K. S. Gibson and J. H. Keegan, J. Opt. Soc. Am. **28**, 372–85 (1936).
[4] J. L. Michaelson, J. Opt. Soc. Am. **28**, 365–71 (1938).
[5] J. M. Moran, Rev. Sci. Inst. **14**, 287–93 (1943).
[6] J. A. Van den Akker, J. Opt. Soc. Am. **33**, 257–59 (1943).
[7] M. Mouton, Ann. Chim. Phys. [5] **18**, 145–89 (1879).
[8] M. E. Carvallo, Ann. Chim. Phys. **4**, 1 (1895); J. Phys. **2**, 27 (1893).
[9] R. V. Baud and W. D. Wright, J. Opt. Soc. Am. **20**, 381–95 (1930).

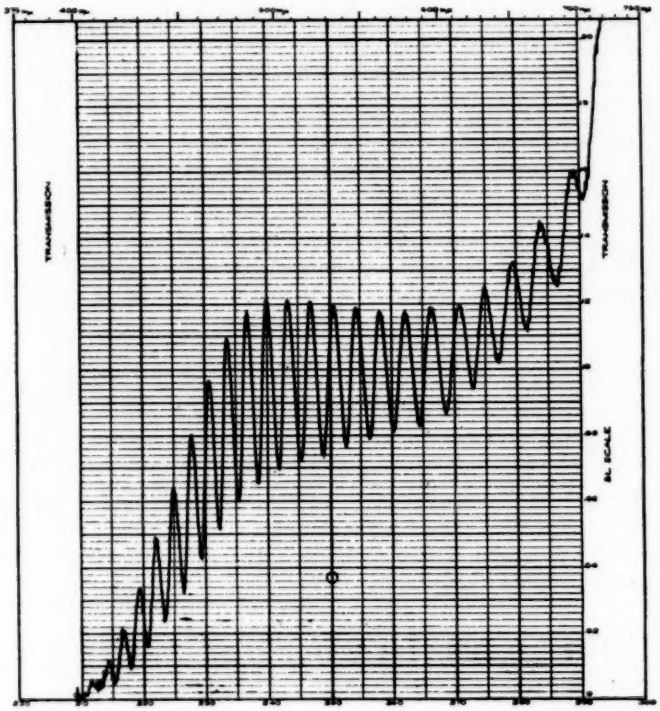

FIG. 1. Transmission curve of a Polaroid-quartz calibrator. The distance between two absorption maximums may be read in millimicrons from the abscissa scale. Expressed in radians, it is 2π in every case.

where D = retardation angle; H = slit width; and $2b/H$ = slit-shape factor.

The result of the above five effects, if absorption is non-selective, is to raise or lower the plotted transmission, but the wave-length position of the points of interest, namely the absorption maximums, minimums, and average transmission, remains unchanged. The following discussion is predicated on the assumption that these effects are not present. This does not limit the usefulness of the method because the plotted curve may be corrected for their influence.

In the visual range a combination of crossed Polaroids with a quartz retardation plate between them may be used as a calibrator. This has been done as shown in Fig. 1. The slight displacements of the observed absorption maximums from the positions predicted with Eq. (1) because of the chromatic absorption of the Polaroids can be calculated in the usual manner for damped sine waves.

Equation (1) was derived on the assumption that the quartz plate of the calibrator was at $\pi/4$. If the quartz plate is not exactly at $\pi/4$ to the polarizer and analyzer, Eq. (1) assumes the more general form

$$T=\sin 2A \sin 2B\left(\frac{1}{4}+\frac{1}{2}\cos D \times\frac{\cos(b-H)-\cos b}{H(H-2b)}\right)+\cos^2(A-B), \quad (2)$$

in which A is the angle between the polarizer and the optic axis of the retardation plate and B is the angle between the analyzer and the optic axis of the retardation plate. Equation (2) is obtained by combining Eq. (1) and Eq. (5) of Part I according to Eq. (4) of Part I.

The transmission is changed by this different positioning, and slit calibration might be affected, but the wave-length positions of the

absorption maximums and minimums are not affected since $dT/dD=0$ at the same values of D independent of A and B. The wave-length position of the maximum is thus not affected by slight errors in positioning the plate. The most useful points are the absorption and transmission maximums and the average transmission points. It is shown below how the positions of these points can be calculated for any particular thickness of quartz with the assumption that H is constant in radians.

Let

$$m=t(n_e-n_o)/\lambda, \tag{3}$$

where m = retardation number $=D/2\pi$, t = thickness of quartz plate, n_e = index of refraction of extraordinary ray, n_o = index of refraction of ordinary ray, and λ = wave-length. Calculate m for all wave-lengths at which the refractive indices are known.[10,11] Approximately 20 wave-lengths are known in the visual and about 20 more between 700 and 1400 mμ. These data are shown plotted in Fig. 2 for $t=2.06$ mm. From Eq. (1) T will be maximum when D is equal to $2n\pi$ where n is an integer and T will be minimum when D is equal to $(2n+1)\pi$. T will be 0.25 when D equals $(2n+1)\pi/2$. The question of determining at what wave-length the transmission becomes maximum, minimum, and 0.25 thus becomes the question of determining at what wave-length m becomes n, $(2n+1)/2$, $(2n+1)/4$, respectively. This may be done graphically by reading from Fig.2. When $m=n$ the retardation plate is a full wave plate, and when $m=(2n+1)/2$ the plate is a half-wave plate. When $m=(2n+1)/4$ the plate is a quarter- or three-quarter-wave plate.

For greater accuracy the values of (n_e-n_o) corresponding to a desired value of m may be found by linear interpolation in a table of m *versus* (n_e-n_o) and substituted in Eq. (3) to solve for λ.

We have used a Nicol or some other polarizing element in this range. The failure of our sample of Polaroid to polarize in the infra-red is evident from the high transmission values of Fig. 1 at the red end.

[10] *International Critical Tables* (McGraw-Hill Book Company, Inc., New York, 1929), Vol. 6, p. 341.

[11] Landolt and Bornstein, *Physikalisch-chemische Tabellen* (Verlagsbuchhandlung, Julius Springer, Berlin, 1923), fifth edition, Vol. II, Table 167.

It is convenient to use a didymium filter to identify rapidly a few of the absorption or transmission maximums as the various calibrator peaks are rather similar.

At any desired wave-length the optimum thickness of plate for wave-length calibration depends only slightly on slit shape. If the half-transmission points are to be used for wave-length calibration the transmission curve with the maximum slope at the half transmission point is the condition for optimum accuracy. If the absorption or transmission maximums are to be used then the curve with the steepest slope adjacent to the maximum will define the wave-length position of the maximum most exactly. The problem in either case thus becomes one of finding for what slit width in radians the slope of the transmission curve of the calibrator is a maximum.

Assuming a rectangular slit, it can be deduced from Eq. (6) of Part I that

$$T=\frac{1}{4}-\frac{1}{4}\cos D\frac{\sin(H/2)}{H/2}. \tag{4}$$

The slope of the curve is the first derivative dT/dD and to find its maximum as a function of H set $d^2T/dDdH=0$. The maximum slope occurs when $H=1.43\pi$. The same value of H maximizes the slope in the case of a symmetrical triangular slit. Intermediate trapezoidal shapes require very nearly the same optimum thickness.

Since a change in D of 2π corresponds to a change of unity in the value of m in Eq. (3),

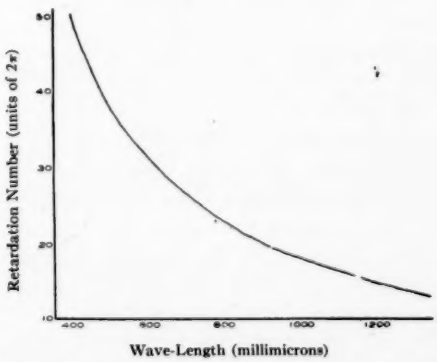

FIG. 2. Relation of retardation number to wave-length calculated from $D=2\pi t(n_e-n_o)/\lambda$ for quartz plate of $t=2.06$ mm.

assuming no change in the refractive indices over the slit width, we may write

$$m = t(n_e - n_o)/\lambda_1, \tag{5}$$

$$m+1 = t(n_e - n_o)/\lambda_2. \tag{6}$$

Subtracting (5) from (6)

$$1 = t(n_e - n_o)[(\lambda_2 - \lambda_1)/(\lambda_1)(\lambda_2)]. \tag{7}$$

Therefore to obtain the best conditions for calibration $\lambda_2 - \lambda_1$ should be $2\pi H'/1.43$ where H' is the slit width in millimicrons. Assuming $\lambda^2 = \lambda_1\lambda_2$ Eq. (7) can be solved for t_0 the optimum thickness in mm.

$$t_0 = 7.15\lambda^2\, 10^{-7}/(n_e - n_o)H'. \tag{8}$$

For quartz $n_e - n_o$ may be taken as 0.009 for this calculation in the visual range.

For example in calibrating in the 550 mμ region a thickness of quartz of about 2 mm is satisfactory when the slit width is 10 mμ. Since $n_e - n_o$ for calcite is about 20 times as large as for quartz the optimum thickness for calcite is about 0.1 mm which makes it less convenient to use in calibration.

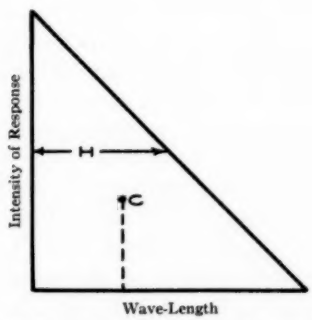

FIG. 3. Geometry of an asymmetrical slit: H is half-height width, C is center of gravity; wave-length position of slit is assumed as being at its center of gravity.

The above discussion has shown how the exact position of the absorption maximums may be determined and has discussed symmetrical slits, the usual condition for narrow slit instruments. It will be noted from Eq. (2) that the position of the plotted absorption maximums, minimums, and average transmissions is independent of slit width, slit shape, and exact angles of optics.

The more general case involves asymmetric slits and this means that the plotted maximums, minimums, and average transmission points will shift with slit width. In evaluating this shift, an extreme asymmetric slit was chosen, having the distribution indicated in Fig. 3. The shift is defined as follows. If two slits were available, one symmetrical and one with the asymmetry of Fig. 3, but with the center of gravity of each slit falling at the same wave-length, and each slit having the same width, then the plotted position of the absorption maximums would not be the same. The difference between plotted positions for the symmetrical and asymmetrical slits is defined as the calibration shift since with a symmetrical slit the absorption maximum is plotted at the true wave-length because the absorption curve is also symmetrical. The following equations apply to the asymmetric slit of Fig. 3 and a calibrator as may be derived with the aid of Eqs. (3) and (4) of Part I.

$$T_a = \frac{1}{2} + \frac{1}{2H}\sin\tan^{-1}\frac{\sin 2H - 2H}{2\sin^2 H} - \frac{\sin H}{2H^2}\sin\left(\tan^{-1}\frac{(\sin 2H - 2H)}{(2\sin^2 \mathrm{H})} + H\right), \tag{9}$$

$$S_a = \tan^{-1}\frac{\sin 2H - 2H}{2\sin^2 H}\cdot\frac{2}{3} - H, \tag{10}$$

where T_a = transmission of the absorption maximum, and S_a = shift of the absorption maximum. The shift in the transmission maximum and in the average transmission points can be shown to be the same as for the absorption maximum.

The correct calibration, as defined above, is one for which straight line absorption curves plot with their transmissions equal to monochromatic values. The calibrator curve at the average transmission point is sufficiently nearly a straight line for slit widths of $\pi/2$ radians or less. Calculation shows that the error due to the slit shape of Fig. 3 in the case of a slit width of $\pi/2$ radians will be 0.64 percent of the slit width. Therefore if the point on the wave-length scale where the average transmission point of the calibrator falls is marked as the wave-length calculated from Eq. (3) with $m = (2n+1)/4$ the calibration error will be 0.64 percent of slit width in this case.

This is essentially Van den Akker's method

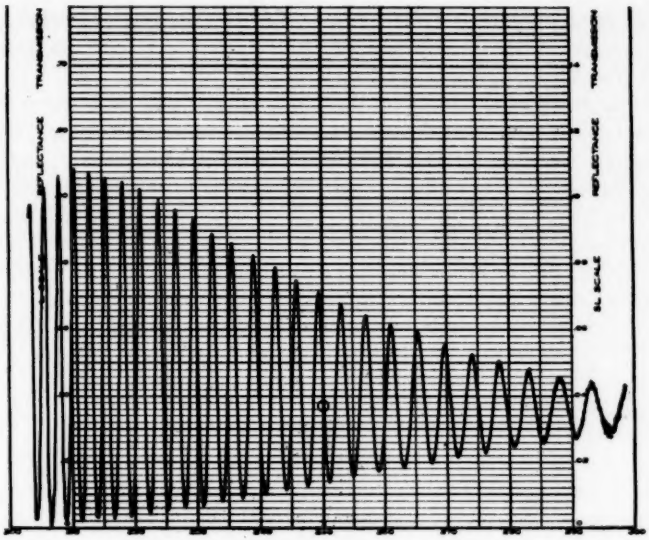

FIG. 4. Calibrator in special case of Martens photometer.

except that the calibrator replaces his filters and the calculation replaces his mercury line determination. This method has the additional advantages of an infinite choice of wave-lengths for calibration and steeper slopes than his filters. Also, the absorption maximums and minimums of the calibrator are subject to asymmetric shifts of the same magnitude as the average transmission points so these may be used equally well as calibration points.

To be sure that the slit width is less than $\pi/2$ radians and hence that the straight-line approximation is valid, the transmission of the absorption maximum should be less than 26 percent of the average transmission. This is calculated from Eq. (9) by letting $H=\pi/2$ radians.

The recommended calibration procedure is then (1) Choose a calibrator of such a thickness that the absorption maximum will be less than 26 percent of the average transmission. (2) Measure the calibrator as a transmission sample. (3) At the points of maximum, minimum, and average absorption mark the wave-length scale with the values calculated by Eq. (3)

The choice of plate thickness so that the transmission minimum will be less than 26 percent of the average transmission assures a satisfactory plate thickness regardless of losses of light owing to surface reflectance and absorption in the calibrator. This procedure will give calibration points at intervals equal to the slit width. If additional points are desired, a second calibrator greater in thickness by 100 $H'/2\lambda$ percent will give an additional and equal number of intermediate calibration points. Any change in thickness, such as the above, may be calculated from Eq. (3).

The limitation of didymium filters as accurate calibrators because of peak asymmetry is generally recognized.[3] The mercury arc calibration neglects the effects of asymmetric distribution of continuous light source, photo-cell response, and absorption of optical parts and may, therefore, be inaccurate. The combination of arc and filters is accurate but limited to the few available emission lines. The difficulties of wave-length calibration in the near infra-red without using the present method have been noted.[1]

For daily calibration the didymium or other permanent filters are convenient but do not have as many calibration points. The calibrator is therefore recommended both for absolute calibration and for daily calibration.

The accuracy of the calibrator is limited, as may be seen from Eq. (3), by measurements of

T and n_e-n_o, and the parallelism of the faces of the retardation plate. Since interferometer measurements will give a very precise result for t and can check the parallelism of the faces, the limiting error is n_e-n_o. These latter data as published at present are cited as accurate to 0.1 part per thousand which corresponds to an uncertainty of 0.05 mμ at 550 mμ. The refractive index differences may be determined with greater accuracy by determining the transmission of a calibrator to emission lines, having the plate thickness measured with an interferometer. The effect of temperature at 508 mμ is to change the calculated positions 1 mμ for a 18°C temperature change.

A special case of interest is a spectrophotometer employing a Martens photometer. Here the retardation plate may advantageously be placed between the Wollaston and the photometering Rochon and fastened relative to the Wollaston with its crystal axis at $\pi/8$ radians to the axis of the spectrophotometer optic which gives a curve as shown in Fig. 4. For the symmetrical slit case it was shown in Eq. (14) of Part I that

$$T=\tan^2\left[-B+\frac{1}{2}\tan^{-1}\left(\frac{H(H-2b)\cot 2B}{2\cos D(\cos(b-H)-\cos b)}\right)\right], \tag{11}$$

where B is the angle of orientation of the retardation plate. From this it is evident that with a symmetrical slit the position of the absorption maximum is independent of slit shape, slit width, and exact experimental angle B.

The wave-length position of the absorption maximum and minimum are calculable in the same manner as the general case. The average transmission is replaced by the average photometering Rochon travel. This means that 17.1 percent transmission marks the $m=(2n+1)/4$ or quarter-wave and three-quarter-wave plate positions since substitution of $D=(2n+1)/2$ in Eq. (11) gives $T=0.171$.

In the case of an asymmetric slit the following equation may be shown to apply when $B=\pi/8$.

$$T_1=\tan^2\left[\frac{-\pi}{8}+\frac{1}{2}\tan^{-1}\frac{H^2}{H\sin\tan^{-1}\dfrac{\sin 2H-2H}{2\sin^2 H}-\sin H\sin\left(H+\tan^{-1}\dfrac{\sin 2H-2H}{2\sin^2 H}\right)}\right], \tag{12}$$

in which T_1 is the transmission in the absorption minimum.

Where the slit width is $\pi/2$, T_1 becomes 74 percent. Problems of absorption and surface reflectance are avoided since the retardation plate affects both beams equally.

The recommended calibration procedure in this case is (1) use a retardation plate of such a thickness that the transmission maximums are 74 percent or above, (2) for each wave-length adjust the photometering Rochon until the two beams are of equal intensity, and (3) at the points of maximum, minimum, and average photometering Rochon rotation, mark the wave-length scale with the values calculated by Eq. (3).

Direct Intensity Measurements of Spectrum Lines with Photo-Multiplier Tubes*

G. H. DIEKE AND H. M. CROSSWHITE
Johns Hopkins University, Baltimore, Maryland

(April 13, 1945)

INTENSITY measurements of spectrum lines have become of increasing importance for scientific problems and industrial applications. The photographic method has been the standard method[1] for many years. It has, however, so many drawbacks, that serious attempts should be made to replace it by another method. Direct photoelectric measurements seem at present to provide the only method that is feasible for anything except the strongest lines.

The recent introduction of photo-multiplier tubes has made it possible to extend photoelectric measurements to relatively weak lines. Several workers have explored this field and shown that with the multiplier tubes the photoelectric method shows great promise. Among these Rank, Pfister, Coleman,[2] and Grimm[3] have secured particularly valuable results.

Some of the chief drawbacks of the photographic method are the necessity for individual plate calibration, its slowness and its limited accuracy because of plate grain, non-uniformity of plates, and erratic behavior due to such phenomena as the Eberhard effect. These difficulties are serious even if, as in spectrographic analysis, only reproducibility is required. They increase if actual intensity ratios are needed, particularly when the lines to be compared are far apart in the spectrum.

Any method with valid claims to replace the photographic method should be free from these difficulties and furthermore should not introduce new complications. For this reason we have investigated in detail the use of photo-multiplier tubes to determine whether they are capable of giving reliable results, particularly for routine applications. We have tried to make use of equipment that is easily obtained and easily handled. It must at the same time satisfy high demands on stability and sensitivity.

The use of photo-multiplier tubes is at present being explored in several laboratories, and recently Hasler and Dietert[4] have described the method used in some of these laboratories. We have thought it advisable to proceed in a somewhat different direction, partly because the method described by Hasler and Dietert is adequately taken care of, partly so that the problem can be looked at from as many different angles as possible, and finally because for our particular purpose another method was more suitable.

While in general it is advantageous to design spectrograph and recording unit together to get the maximum efficiency, the system we have employed can be installed without much trouble on almost any kind of spectrograph. Most of our results were obtained with a 21-foot grating in a Wadsworth mounting. The equipment was designed for maximum versatility. It can be used equally well for photography and direct photoelectric readings, and can be converted from one mode of operation to the other within a few seconds.

The multiplier photo-tube is mounted behind a slit on a little carriage so that this slit, the exit slit, can move along the exact focal curve of the spectrograph. This is achieved by attaching the tube carriage to a steel tape which is moved uniformly by a synchronous motor. The tension of the tape presses the carriage against the rails defining the focal curve. The tape moves the exit slit, with the tube behind it, and thus scans the spectrum. The rails are brought in exact coincidence with the focal curve photographically, and this can be done within a few minutes, so that no time is lost when changing to a different wavelength region. If desired several tube carriages can be used simultaneously. The scanning speed

* Results of work on a project sponsored by the Office of Production Research and Development of the War Production Board.

[1] For a review, see G. R. Harrison, J. Opt. Soc. Am. **19**, 277 (1929).

[2] D. H. Rank, R. J. Pfister, and P. D. Coleman, J. Opt. Soc. Am. **32**, 390 (1942).

[3] D. H. Rank, R. J. Pfister, and H. H. Grimm, J. Opt. Soc. Am. **33**, 31 (1943).

[4] M. F. Hasler and H. W. Dietert, J. Opt. Soc. Am. **34**, 751 (1944).

can be adjusted to the proper value by a suitable choice of gears.

Obviously the tubes can also be used in a stationary position. In general, however, scanning has definite advantages. With stationary tubes, unless very wide slits are used, the position of the exit slit is very critical. Slight changes in the set-up, for instance simply a change of the room temperature, may affect the adjustments. With the moving slit all that is necessary is to bring it somewhere in the vicinity of a line. With stationary slits it is also difficult or impossible to obtain lines lying close together, which presents no difficulties whatsoever with a moving slit. Different parts of the photo-cathode become illuminated as the slit moves over a large part of the spectrum. This is so since the focal curve is not a circle with the grating as center. These different parts may have varying sensitivities. Unless the image actually moves off the cathode, this does no harm. To obtain true intensity ratios over a wide wave-length interval, the instrument must be calibrated anyway, and this shift is taken care of by the calibration. If it is necessary to install this device on a short focus prism or grating spectrograph where the shift is too large, another degree of freedom must be provided for the motion of the tube in order to keep the image always centered.

The multiplier tubes need a power supply furnishing about 1200 volts d.c. They are very sensitive to even small changes in the voltage. We found that by a combination of a constant voltage transformer and nine voltage regulator tubes, no difficulties were experienced when the power supply was taken from a rather unsteady 110-volt city a.c. line. Tests showed that a variation in the line voltage from 90 to 120 volts could not be detected in the output of the photo-tubes.

The photo-current can be measured with any microammeter or galvanometer of sufficiently high sensitivity. A recording instrument is particularly useful. We are indebted to Mr. I. N. Stein, Dr. P. H. Dike, and Mr. A. J. Williams for the use of an electronic recording microammeter recently developed by Leeds and Northrup.[5] It has a full scale deflection of 5 μA and can be used for higher currents by suitable shunts. It has worked dependably now for many months. With reasonable shielding a powerful spark source can be operated in the immediate vicinity of the spectrograph without affecting the photo-current.

For such light sources as are used in industrial spectrographic analysis, the sensitivity of this instrument is ample. If for special purposes lower intensities need to be measured, the current is amplified first in a d.c. amplifier. We used amplifications up to 400 times, but generally there is no need to go beyond 200 times, because the fluctuations in the dark current will become troublesome at such high amplifications. For the maximum amplification that is permissible, see Rank, Pfister, and Grimm.[3]

Without the d.c. amplifier the fluctuations in the dark current are unnoticeable.

We believe that an instrument as described here gives results superior in almost all respects to the photographic method. This is true when actual intensity ratios are required and accuracy is necessary, while convenience of operation, although desirable, is of secondary importance, and also for industrial analysis where reliability under routine conditions and ease of operation are just as important as a reasonable accuracy. The great flexibility of the electrical measurements makes it possible to adjust the mode of operation to fit the particular problem, and we have tried out many variants of the procedure described here. A few of the results will illustrate what can be achieved.

Figure 1 shows a part of the iron spectrum. The middle is the reproduction of a photograph, the upper part a microphotometer trace of the photograph, and the lower part a direct photoelectric recording of the spectrum. Notice the grain of the plate on the densitometer tracing and the complete absence of any ripple on the lower part.

Resolving power. What can be ultimately achieved is of course determined by the resolving power of the spectrograph. If the highest resolving power is to be achieved, a certain sacrifice in sensitivity must be accepted and vice versa. As there is an ample reserve of sensitivity this presents no difficulties. We find that without trouble we can actually achieve the same resolving power with photoelectric recording as with

[5] I. M. Stein, A. J. Williams, Jr., and W. R. Clark, Trans. A.S.M.E. (May, 1944).

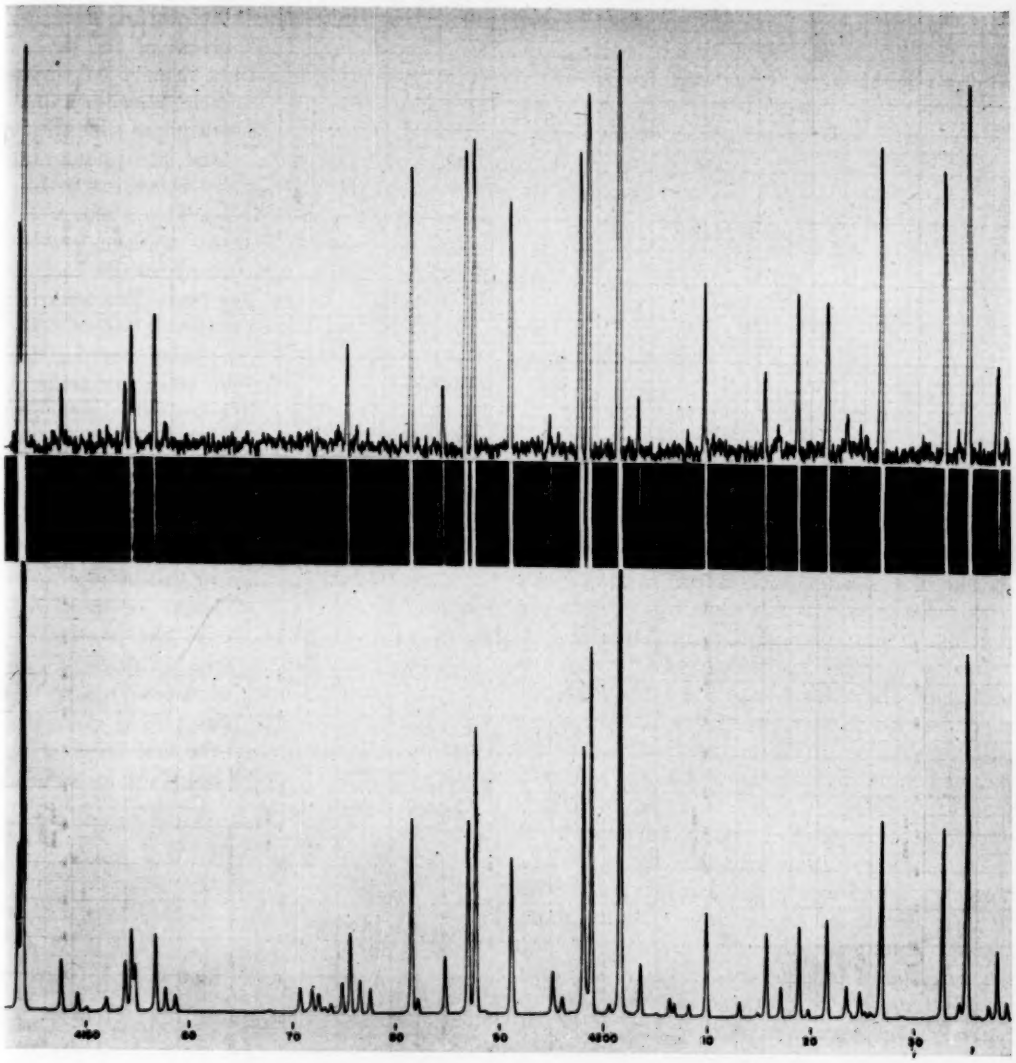

FIG. 1. Part of iron spectrum. *Middle:* reproduction of photograph. *Upper:* microphotometer trace of photograph. *Lower:* direct photoelectric recording.

photography. This is brought out in Fig. 1 and perhaps even better in Fig. 2, which shows a recording of the familiar 3100 triplet of iron (second order). Note that the short wave-length component is resolved, although the two lines are separated by less than 0.07A. Figure 3 shows the 3175 Sn line and the 3175 Fe line taken with a somewhat wider slit.

Sensitivity. Rank, Pfister, and Coleman[2] have given as a good estimate of the attainable sensitivity that anything that can be photographed within two hours can be measured photoelectrically without trouble, and our own experience agrees with this, although of course exact values depend very much on the particular conditions. Figure 4 shows the surroundings of the 4358 mercury lines with different sensitivities. *A* and *B* are without, *C* and *D* with an amplifier.

D has 500 times the sensitivity of A. On D the 4358 line would be 43,000 scale divisions high, while the sixth-order ghost (at about 4390A) is about one scale division. The change-over from scale D to A can be made in a fraction of a second,

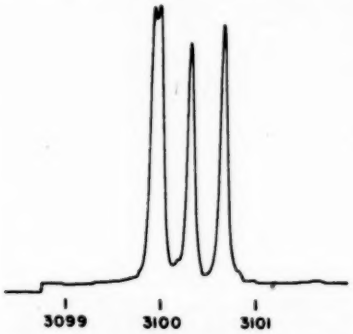

FIG. 2. 3100A triplet of iron (second order).

so that there is no difficulty in measuring intensity ratios as high as 40,000:1, which is quite beyond the range of any photographic plate.

Change of sensitivity with wave-length. Figure 5 shows how the sensitivity of the 931A and IP 21 tubes depends on the wave-length. The curve was obtained by comparing the photo-tubes with a thermopile. It depends somewhat on the individual tube and the way it is used. Figure 6 shows the response in the spectrograph, which depends of course on the particular grating used. This, however, is typical of a great many gratings in common use. This curve is of course the combination of Fig. 5 and the sensitivity curve of the spectrograph. Although these can be determined separately, a direct determination of the resultant Fig. 6 is much easier and more reliable. We have not yet had a chance to work with the IP 28 tube, which is sensitive down to 2000A. However, by measuring the absorption of the glass envelope of the 931A tube and correcting for it, the dotted part of Fig. 5 is obtained, which shows the sensitivity of the 931A tube in an envelope of quartz or glass that does not absorb appreciably above 3000A. We expect to be able to supply soon similar data for the IP 28 tube down to 2000A.

The sensitivity of the photographic plate also changes strongly with wave-length. This is shown for three commonly used types of plates in Fig. 7. The curves are obtained by plotting a quantity proportional to the reciprocal of the exposure time necessary to obtain a density 0.6 and are therefore affected by the reciprocity law failures. Figure 8 shows the ratios of photo-tube to plate sensitivity. This shows that the photographic plates lose their sensitivity faster in the ultraviolet than the photo-tubes. This is also true for longer wave-length, but by selecting panchromatic or infra-red sensitive plates this ratio can be changed in favor of the plate. The sensitivity of the photo-tubes becomes small beyond 6000A. This is due to the particular photo-cathode employed. Other multiplier tubes are made and undoubtedly will soon be available, which are sensitive to 12,000A. We can therefore say that the multiplier tubes can cover efficiently the same wave-length range as the photographic plates.

For use with grating spectrographs, it is important that the instrument can be used in the various orders without danger of interference from overlapping orders. In the photographic procedure this can be achieved in almost all cases by a proper combination of plate and filter. With the present multiplier tubes (931A) there is no difficulty whatsoever for the first order of any wave-length, as filters can always be found which

FIG. 3. Fe 3175 and Sn 3175.

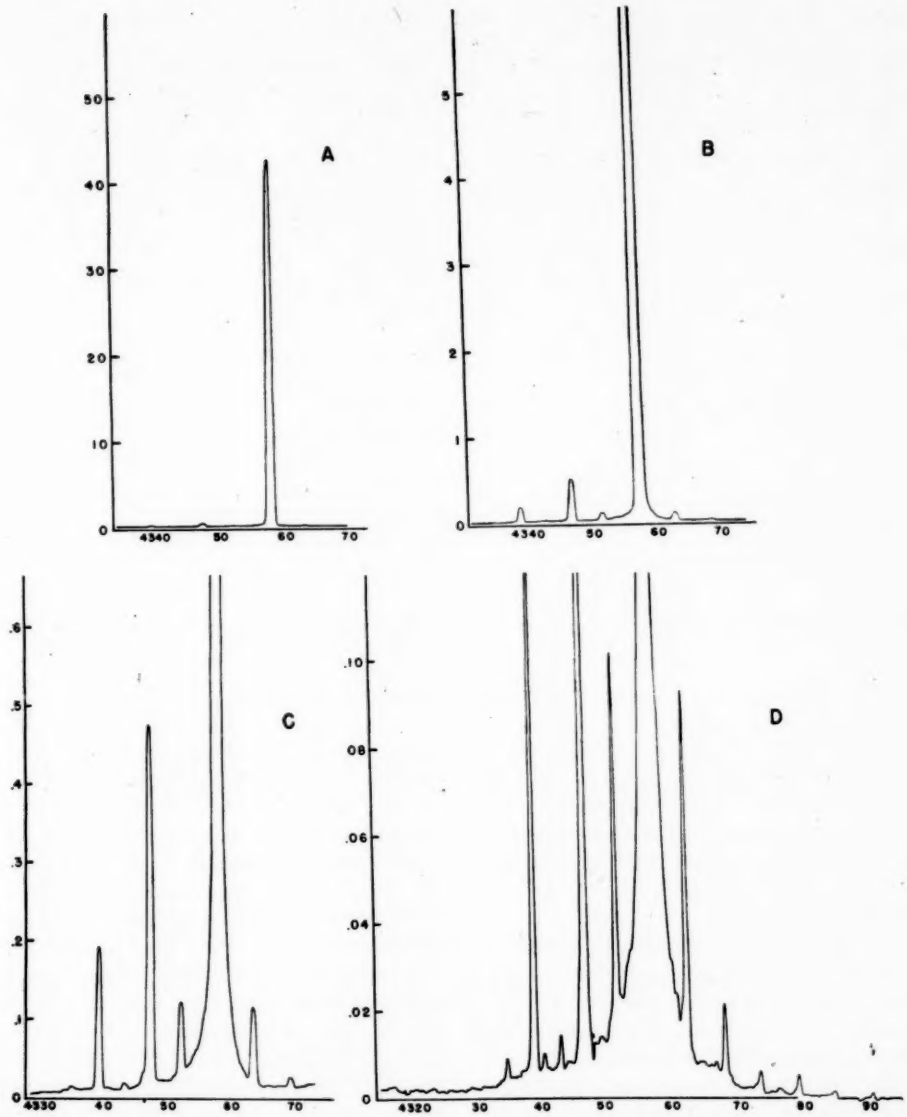

FIG. 4. The 4358 mercury line recorded with different sensitivities. Besides the main line 4358 and its ghosts up to the sixth order, the weaker lines 4348, 4339, and 4344 appear.

cut out effectively the higher orders. There is no obstacle to the use of the second order when filters of the Corex type are used. There may be difficulties at present for some regions of the third or higher orders, but these orders are not ordinarily of great importance. It is expected that the tube manufacturers can provide photo-surfaces which have their sensitivity in a narrower region and thus make it possible to isolate also the higher orders of the grating if that should become necessary.

The linear response of the tubes is good, so that intensity measurements of neighboring lines can be made without difficulties. For actual intensity

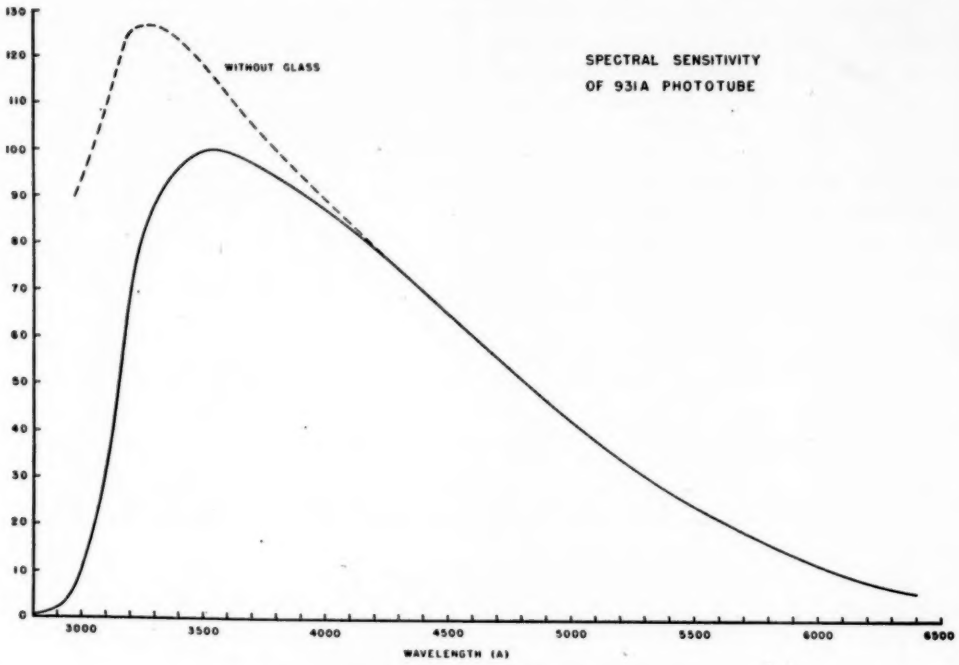

FIG. 5. Tube sensitivity as function of wave-length.

measurements over wide wave-length intervals it is necessary to calibrate the combination: light source, spectrograph, and photo-tube as it is actually used. This can be done readily for wave-lengths longer than 3000A with a calibrated tungsten lamp, or in general with any calibrated reproducible light source.

For spectrographic analysis, where only reproducibility is required, such a calibration is not required if conditions are left unchanged.

Controlling tube properties. The present photomultiplier tubes have highly individualistic properties, and two tubes of the same lot may be very different. This is less true for the selected IP 21 tubes than for the 931A tubes. There are great differences in sensitivity, dark current, and other properties. The photo-cathode may be very non-uniform. All these things must be taken into account but can be compensated for by a proper technique.

Probably the greatest source of potential trouble is the property of the tubes known as *fatigue*. This is a decrease of sensitivity with continued use. The original sensitivity may be regained after a period of idleness, which is usually considerably longer than the period during which the fatigue was acquired. The amount of fatigue is very different in the individual tubes and may be as much as to decrease the sensitivity to 30 percent after an hour's use. The amount of fatigue depends also on the intensity of the light. Quantitative data on fatigue will be found in the complete report mentioned at the end of this paper. Unless guarded against, fatigue may produce quite serious errors. The best policy is not to have the tubes illuminated longer or more strongly than necessary to make the measurements, to avoid the tubes with strong fatigue, and to use matched tubes when more than one tube is necessary.

APPLICATIONS FOR INDUSTRIAL SPECTROSCOPY

Although the particular method of using photomultiplier tubes for spectroscopic intensity measurements described here was first developed to have a reliable method of studying the properties of light sources, its advantages over photographic methods in routine industrial analysis were so

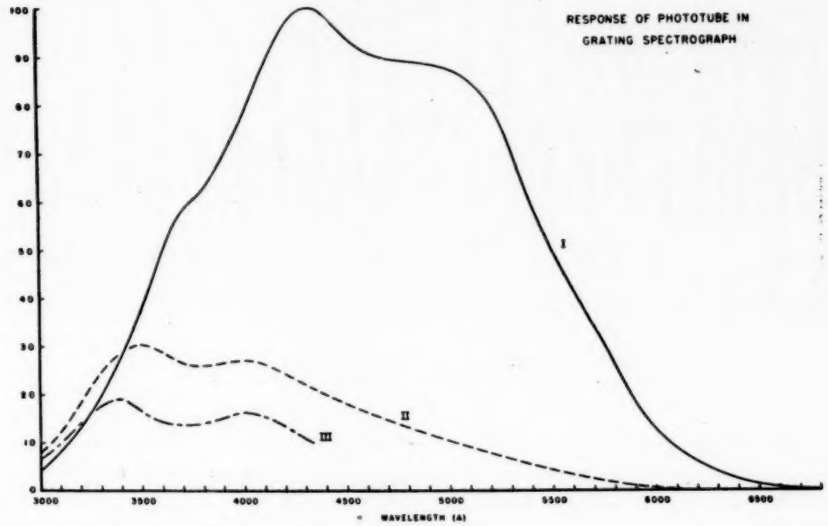

FIG. 6. Response of photo-tube in grating spectrograph. First, second, and third order.

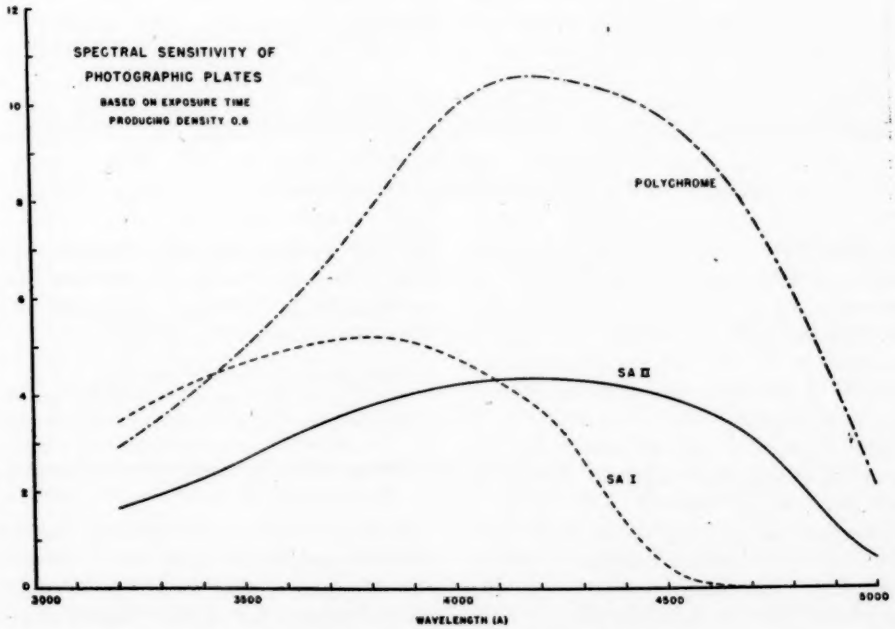

FIG. 7. Spectral sensitivity of three photographic plates.

obvious that its possibilities in that direction were further explored.[6]

[6] See also E. A. Boettner and G. P. Brewington, J. Opt. Soc. Am. **34**, 6 (1944).

It can be specially adapted for each particular problem and in general will be vastly superior, as far as intensity measurements are concerned, to the photographic procedure in accuracy, speed,

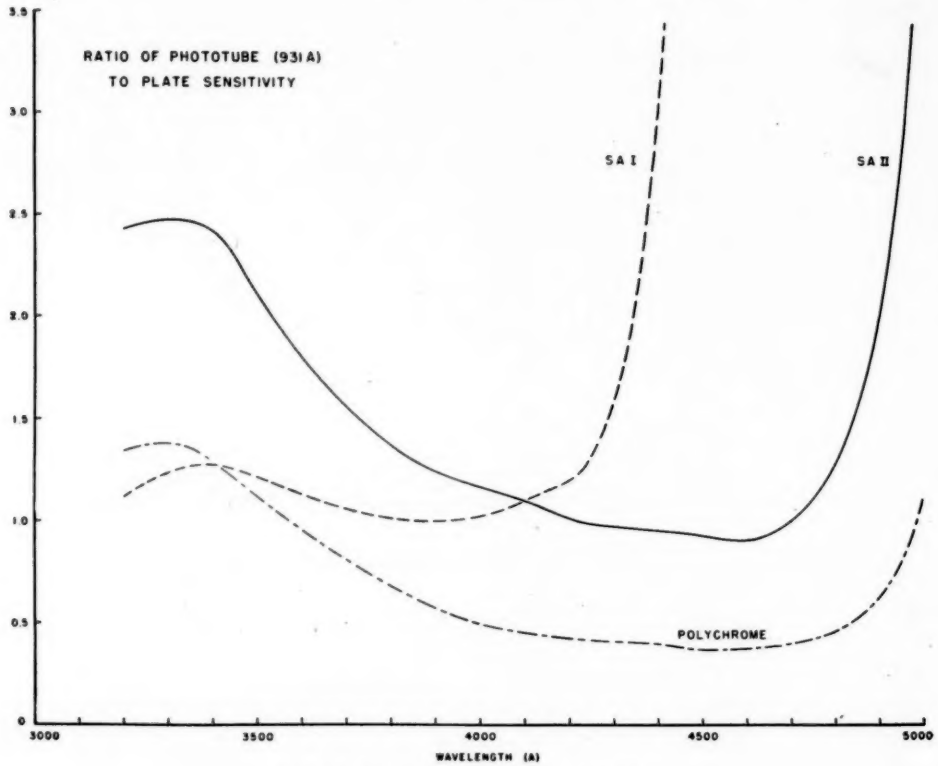

FIG. 8. Ratio of photo-tube to plate sensitivity.

and reliability. An exception is the case when large numbers of lines will have to be observed simultaneously, as for instance in a qualitative analysis. The photoelectric method will find its greatest strength in routine industrial control processes where the same type of measurements will have to be repeated many times.

Although because of the temporary lack of tubes sensitive in regions where most of the lines lie which are usually used for spectrographic analysis, we have not yet had the chance to carry out many of the more familiar analyses, nevertheless we can give a few examples of the particular technique with which we have worked.

Suppose a number of lines are to be measured. If the lines are spaced so far apart that a separate tube can be used for each line, the procedure is particularly simple, and can be so arranged that it works completely automatically. However, the procedure is hardly less simple if some or all of the lines lie close together. The electrodes are inserted and the closing of one switch actuates successively the following operations, which are controlled accurately by a timer:

1. Beginning of the scanning operations.
2. Starting of the light source, at exactly the same time as 1.
3. Switching the recorder successively to each tube just at the time when its slit passes the proper line.
4. Stopping of light source, scanning, and recording, so that the instrument is ready for the next sample.

In order to have a guarantee that the scanning is done in exactly the same way every time, the scanning range is determined by two adjustable limit switches which reverse the direction of the scanning motor when the tube carriage touches the switch. It is convenient to have half the lines recorded scanning toward long wave-lengths and half when returning to shorter wave-lengths. If it is arranged so that the period of the timer controlling the other operations is just a little shorter

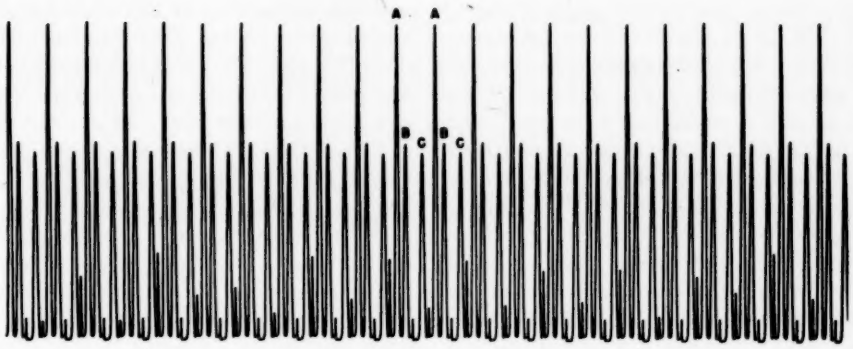

FIG. 9. Automatic record of the relative intensities of three mercury lines. *A*: 4358A; *B*: 5461A; *C*: 4047A.

than the period determined by the limit switches, and the cycle begins when the tube reaches the first limit switch, then all operations occur exactly the same way each cycle. If therefore there is partial evaporation in the light source, each line will always be recorded after exactly the same pre-burning time even though this pre-burning time is different for different lines. In this way reproducible results are assured.

It should be emphasized that none of the adjustments necessary to get the equipment ready for a certain sequence of lines is complicated or delicate. They can be made in a very short time. The approximate settings of the tubes are made with a wave-length scale attached to the focal curve. This means that this method is suitable when the type of work required of the spectrograph is often changed. Once timer and tubes are set, nothing need be touched for an indefinite number of measurements of the same line group.

The working of this arrangement is illustrated in Fig. 9 for the measurement of the three mercury lines 4358, 4047, and 5461. This mercury tube provides a very steady light source so that variations in the light as present in most other sources cannot confuse the results. About five seconds are allowed for every reading. This time, with a steady light source, is determined by the period of the measuring instrument. It might be reduced to two seconds per reading for our recorder, but then the phasing would have to be more accurate.

The instrument was left running unattended for more than two hours with an automatic reading every five seconds. The maximum deviation did not exceed one percent, and that was very probably not due to the measuring device but to changes in the light source. With a somewhat different arrangement we have been able to make measurements to an accuracy of 0.1 percent, although we have not found a light source which stays constant to this extent.

If there are lines so close together that separate tubes cannot be used, the same tube will scan both lines. If they are not immediately adjacent, after recording one line, the instrument may be switched to other tubes and back to the first tube when it comes to the second line. Obviously mechanical arrangements can be made that the tube will jump through a certain wave-length interval during this time. With a proper arrangement, therefore, any group of lines can be taken care of.

If the lines to be measured differ greatly in intensity, it can be so arranged that a shunt of proper size is automatically switched into the circuit for the stronger line so as to give suitable deflections for both lines.

The recorded intensities of the lines in terms of the internal standard can then be converted into concentrations with the help of the working curve. Once the working curve is established, a scale can be constructed with which the concentrations can be read directly from the original chart. Therefore little more than five seconds per determination need be allowed from the time the electrodes are placed in the light source.

Whereas the experiments performed so far show that the problem of measuring intensities

presents no difficulties with the photo-multiplier tubes, the difficulties inherent in the light source are accentuated. For the measurements are made during a very short time, and the full fluctuations of the light source are recorded. In the photographic method the fluctuations are averaged out, and this is also done in the photoelectric method described by Hasler and Dietert. We believe, however, that here the improvements should be made at the light source. Even when averaging over a long period, there is no guarantee that the fluctuations cancel; often this is definitely not so.

We are working at present on this problem in several ways, and the prospects are hopeful. One way of solving the difficulty is by increasing the electrical period of the circuit to such values that integration over an appreciable time occurs. This would be equivalent to the integration over the period of the photographic exposure. This slows down the whole recording process and is therefore less desirable. If necessary the time lost can be recovered by using two or more recorders.

Experiments with measuring circuits for which the relative intensities are recorded independently of intensity fluctuations and with methods of control of the light source so that there are no fluctuations or that their effect on the results is eliminated are still in the preliminary stages.

The results described here are one phase of work carried out at Johns Hopkins University under the sponsorship of the Office of Production Research and Development of the War Production Board. Details will be found in a forthcoming report issued by this office.

Tests for the Detection and Analysis of Color-Blindness. III. The Rabkin Test

LEGRAND H. HARDY, GERTRUDE RAND, AND M. CATHERINE RITTLER
Knapp Memorial Laboratories, Institute of Ophthalmology, College of Physicians and Surgeons of Columbia University, New York, New York

(Received April 19, 1945)

IN a previous report[1] we discussed the general situation with respect to polychromatic test plates for the detection of defective color vision, pointing out the advantages and limitations of this type of test and the need for care in administration. Stress was laid on the following four important sources of error which, aside from defects in the tests themselves, account for part of the failure of the tests to screen out all the color defectives: 1. Disregard of the critical importance of illumination, 2. Careless and inefficient administration, 3. Incompetent and invalid interpretation of results, and 4. Misapplication of the tests to fields for which they were not intended, or the expectation of more from the tests than they are capable of, or designed for, yielding.

In that report a critical evaluation was made of the fifth edition of the Ishihara test for the analysis and detection of defective color vision. In a second report[2] the study was extended to include other representative editions of the Ishihara test. In these reports analysis was made (a) of each plate of the test as a medium for detecting defective color vision, (b) of the *significance of performance scores* on the entire test both for the detection of the presence of color defect and for the equally important classification of type and extent of the defect, and (c) of the *type of response* to certain plates as a clue to classification.

In the present report a similar evaluation is made of Rabkin's Polychromatic Plates for Testing Colour Vision. We find this test to be in many respects much more valuable and informative than the Ishihara. It has equally good possibilities for screening color defectives from color normals and in addition, when scored by the revised procedure to be described, yields, in 93 percent of the red-green defectives tested, correct information as to the nature of the red-green defect present. We do not find it, however, adequate to differentiate between dichromats and anomalous trichromats, as is claimed by Rabkin.

The Rabkin test was first published in 1936.[3] A second edition appeared in 1939.[4] The second edition does not differ from the first in general form. However, some of the pigments have been changed as the result of experimentation, and substitutions have been made for some of the plates. The test is distributed in this country by Four Continent Book Corporation, 253 Fifth Avenue, New York, New York.

The Rabkin test is interesting in that it was designed to separate anomalous trichromasy from dichromasy and at the same time to furnish a differential diagnosis as to type of dichromasy (protanopia, deuteranopia, tritanopia) as well as to type of anomalous trichromasy (protanomaly, deuteranomaly). It is claimed and, so far as we know, justly, that this is the first attempt to separate anomalous trichromasy from dichromasy by means of polychromatic printed plates. The subdivision of dichromasy suggested by von Kries into protanopia and deuteranopia (1897)[5] and tritanopia (1911)[6] and of trichromasy into normal and anomalous suggested by Nagel (1904)[7] is followed by Rabkin. He does not, however, recognize the subdivision of blue-yellow blindness proposed by G. E. Müller (1924)[8] into

[1] L. H. Hardy, G. Rand, and M. C. Rittler, "Tests for the Detection and Analysis of Color-Blindness. I. The Ishihara Test: An Evaluation," J. Opt. Soc. Am. **35**, 268 (1945).

[2] L. H. Hardy, G. Rand, and M. C. Rittler, "Tests for the Detection and Analysis of Color-Blindness. II. The Ishihara Test: A Comparison of Editions," J. Opt. Soc. Am. **35**, 350 (1945).

[3] E. B. Rabkin, *Polychromatic Plates for Color Sense Examination* (Kief and Kharkof, State Medical Publishing Board, 1936).

[4] E. B. Rabkin, *Polychromatic Plates for Testing Colour Vision* (State Medical Publishing House of the USSR, 1939), second edition.

[5] J. v. Kries, "Ueber Farbensysteme," Zeits. f. Psychol. u. Physiol. d. Sinn. **13**, 241–324 (1897).

[6] J. v. Kries, "Normale und anomale Farbensysteme" in H. v. Helmholtz, *Handbuch der physiologischen Optik* (L. Voss, Hamburg u. Leipzig, 1911) third edition, Vol. 2, pp. 333–347.

[7] W. A. Nagel, "Die Diagnose der anomalen trichromatischen Systeme," Kl. Monats. Augenheilk. [1]**42**, 366–370 (1904).

[8] G. E. Müller, *Darstellung und Erklärung der verschiedenen Typen der Farbenblindheit*, u.s.w. (Vandenhoed und Rupprecht, Göttingen, 1924).

tritanopia (violet-yellowish green defect) and tetartanopia (blue-yellow defect). Rabkin defines tritanopia as "blue or violet colour-blindness." He states: "In cases of tritanopia the spectrum is shortened at its violet end, the maximum of brightness is at the sodium line (line Na), and there are two achromatic zones—one in the blue colour, the other in the yellow colour regions."

Theoretically, the method used by Rabkin in designing his plates was based on the fact of "different perception of the hues in the warm part of the spectrum for the normal trichromats and the dichromats and also on the differences of the brightness distributions in the spectrum for various colour systems" (second edition, p. 31). That is, in planning plates designed to screen dichromats from trichromats he played mainly on the aspects of color confusion in the red-yellow-green region and on the presence of a neutral zone in the dichromatic spectrum; and in planning plates designed to distinguish deuteranopic and protanopic color vision, he utilized chiefly the characteristic shift in the brightness of the colors and the different location of this neutral zone. For deuteranopes, he locates the neutral zone in the green at approximately 500 mμ, and for protanopes, in the blue-green at approximately 490 mμ. He does not mention or utilize the second neutral region in the color scheme for these dichromats (in the red-purple region for deuteranopes and the red for protanopes). This, we believe, is unfortunate since it would have increased the value of some of the plates for differentiating between deuteranopia and protanopia. For the differentiation between anomalous trichromasy and dichromasy, he relies on the correct reading of certain specified plates.

The series comprises 20 plates in most of which digits or simple geometrical forms composed of colored disks varying in size, chroma, and value (saturation and brightness) appear on backgrounds composed of colored disks similarly varying in size, chroma, and value. In certain of the plates all of these variants are not used. Plates Nos. 1 and 2 are mainly for purposes of demonstration; Plates 3 through 12, and Plate 15 are intended for screening color defectives from color normals; Plates 13, 14, 16, 17, and 18 (in all of which more than one combination of figure and background appear) are used for differentiating between deuteranopes and protanopes; and Plates 19 and 20 are used for the detection of tritanopia. The correct reading of one or several of Plates 3, 4, 5, 6, 11, 16, 17, and 18 is the indication of anomalous trichromasy when the other plates are read the same as in dichromasy. The type of anomaly (deuteranomaly or protanomaly) is determined by the occurrence of typical deuteranopic or protanopic responses or, in case a diagnostic plate is correctly read, by ascertaining which of the two digits or figures is seen more clearly. The instructions on this interpretation of the responses are rather detailed and require careful study to do justice to the diagnostic possibilities of the test.

Because the Rabkin test is not widely known or easily accessible, it may be worth while to describe some of the more unusual plates in detail. Plates 3, 4, 5, 9, and 11 are of the "transformation" type used by Ishihara, where the color normal sees one figure and the color defective another. These plates are by no means a copy of the Ishihara plates and the figure seen by the color defective is much less visible to color normals than it is in the Ishihara plates. Plates 6, 7, 8, and 10 are of "the vanishing pattern" type used by Stilling, Ishihara, and others, where the color normal sees a pattern and the color defective does not. In the Rabkin plates of this type the patterns vary as to visibility. The digit of Plate 8, for example, is not seen by many persons of low normal color sensitivity. In Plates 13, 14, 15, 16, and 17, different combinations of figure and background are ingeniously presented in various segments of the plate, some of which are seen by deuteranopes, some by protanopes, some are transformation patterns, and some are hidden figures not seen by color normals but visible to most red-green defectives. These are the plates chiefly used in classifying deuteranopes and protanopes. A table of typical responses provides the key to diagnosis. Plate 18 is of particular interest as a test to differentiate between anomalous trichromasy and dichromasy. It consists of small red and green squares arranged on a gray background in a square formation so that there are 8 vertical arrays, numbered 1–8, and 8 horizontal arrays, numbered 9–16. Rows 9, 11, 13, and 14 consist of red squares, some darker than others; and rows 10, 12, 15, and 16 of green

squares, some lighter and bluer than others. Columns 1, 2, 4, 6, and 8 contain the red and green squares of the same brightness, and columns 3, 5, and 7 contain the red and bluish green squares of unequal brightness. Instructions are to select those arrays which are one-colored, although there may be variation in brightness. Color normals are supposed to select horizontal arrays 9–16; deuteranopes, vertical arrays 1, 2, 4, 6, and 8; and protanopes (who see red as darker and green as lighter than do normals and deuteranopes), vertical arrays 3, 5, and 7. Anomalous trichromats are supposed to select either horizontal arrays alone, or some horizontal and some vertical arrays. In the latter case (mixed arrays) the color defect is more extreme and the diagnosis of anomaly is clean-cut; in the former case (horizontal arrays only) the color defect is less pronounced and the diagnosis that an anomaly is present is based on the failure to read certain other plates of the series.

Rabkin does not consider performance scores on the entire test, therefore he does not give any critical score to indicate the dividing line between low, but normal color vision and defective color vision.

EVALUATION OF THE SECOND EDITION OF THE RABKIN TEST AS A MEANS OF DETECTING AND CLASSIFYING DEFECTIVE COLOR VISION

The Rabkin Test (second edition) was included as one test in a battery used in the Knapp Memorial Laboratories to study the cases of defective color vision referred to the laboratories. These cases were not selected on the basis of any statistical survey. They included patients referred to us by ophthalmologists, men who had been rejected by the Armed Forces on the basis of color defect, cases obtained from some of the New York City High Schools, from Vanderbilt Clinic, etc. The battery of tests included the following which have been described in the literature:

Ishihara Tests for Colour-Blindness (fifth edition)[9]
Rabkin Polychromatic Plates for Testing Colour Vision (second edition)[4]
Pseudo-isochromatic Plates for Testing Color Perception engraved and printed by Beck Engraving Company, Inc., and published and distributed by the American Optical Company[10]
The Farnsworth-Munsell 100-Hue Test[11]
Farnsworth Dichotomy Test[11]
I.S.C.C. Single Judgment Test for Red-Green Discrimination[12]
Nagel Anomaloscope[13]

Several additional tests were used which have not as yet been described in print, notably a series of Polychromatic Plates and a new form of Anomaloscope devised in this laboratory. The results of the entire battery of tests aided in the classification of the color defect as to type and extent. Our data thus permit us to evaluate each of the above tests as a screening medium and as a medium for differential classification. The present paper, however, will be confined to the Rabkin test, mention being made of the battery of tests because of its use in classifying the subjects as to type and extent of the color defect.

The classification of the subjects followed the scheme outlined by two of us in "Recent Developments in Color Vision Testing."[14] To those not familiar with the classical terminology of color-blindness and to whom this reference is not available, the following simple explanation of terms may be of value.

On the Young-Helmholtz theory there are *3 factors in the color vision process:* 1. The red primary, *proton*, or first process; 2. The green primary, *deuteron*, or second process; and 3, the blue primary, *triton*, or third process. Hence we should expect to find *3 types of color vision.* Trichromatic, in which all 3 processes function; dichromatic, in which only two processes function; and monochromatic, in which there is only one differentiated physiological process. Trichromatic color vision means all 3 processes are functioning, but since one or more of the processes may function aberrantly there will be as a result *3 types of trichromatic color vision:*

[9] S. Ishihara, *Tests for Colour-Blindness* (Tokyo, Kanehara, 1925), fifth edition.

[10] American Optical Company, *Pseudo-Isochromatic Plates for Testing Color Perception* (Beck Engraving Company, Philadelphia and New York, 1940).

[11] D. Farnsworth, "The Farnsworth-Munsell 100-Hue and Dichotomous Tests for Color Vision," J. Opt. Soc. Am. **33**, 568–578 (1943).

[12] L. H. Hardy, "A Single Judgment Test for Red-Green Discrimination," J. Opt. Soc. Am. **33**, 512–514 (1943).

[13] W. A. Nagel, "Zwei Apparate für die augenärztliche Funktionsprüfung. Adaptometer und kleines Spektralphotometer (Anomaloskop)," Zeits. f. Augenheilk. **17**, 201–222 (1907).

[14] L. H. Hardy, and G. Rand, "Recent developments in color vision testing," Am. Acad. Ophth., Outlines of Instructional Courses, 1944, Graduate Lecture, Continuous course No. 6 and 7.

1. Normal; 2. Anomalous, in which one or more of the processes is weak; and 3. Low discrimination in which apparently all processes are weak. The Anomalous Trichromats are designated according to which process is weak as: 1. Protanomalous, or predominantly red weak; 2. Deuteranomalous or predominantly green weak; and 3. Tritanomalous or predominantly blue-yellow weak. Dichromatic color vision means that only two of the processes are functioning and hence is similarly divided into *3 types:* 1. Protanopia, formerly called red blind, 2. Deuteranopia, formerly called green blind, and 3. Tritanopia, formerly called blue-yellow blind. As a mnemonic it may be remembered that there are *3 kinds of abnormal color vision:* 1. Anomalous Trichromatism (*3 types:* protanomalous, deuteranomalous, tritanomalous), 2. Dichromatism (*also 3 types:* protanopia, deuteranopia, tritanopia), and 3. Monochromatism. Protanopia exhibits a shortened red end of the spectrum, a neutral area around 493 mμ (in the blue-green), a second neutral band in the red (at 493 c), and a brightness peak which is shifted toward the violet. Deuteranopia shows a neutral region in the green (about 497 mμ), a second neutral band in the red-purple (at 497 c), and no shift in the luminosity peak. Tritanopia shows a shortened blue end of the spectrum and a neutral band in the yellow-green. The luminosity peak is not shifted. These types, as well as corresponding anomalous trichromatic types, are best detected and classified by anomaloscope tests.

The subjects whose responses are used for the evaluation of the Rabkin test numbered 106. On the basis of their responses to the entire battery of tests, 74 were classified as having defective color vision of the following types: 35 were anomalous trichromats of which 23 were deuteranomalous and 12 protanomalous; 32 were dichromats of which 15 were deuteranopes and 17 protanopes; and 7 showed generally low color discrimination (that is, low discrimination to all colors) sufficient to amount to a color defect. The remaining 32 subjects showed low color discrimination to a lesser degree which would not be sufficient to amount, in all probability, to a real color defect. This last group we have called "low normals." In order to establish the presence of dichromasy and anomalous trichromasy, particular reliance had to be placed on the results obtained with the Anomaloscope, the Farnsworth Dichotomy test, and our own Polychromatic Plates; and in order to establish the presence of low color discrimination and low normal color vision, on these tests and in addition the Farnsworth-Munsell 100-hue test.

We recognize of course that the dividing line between the best of the low discrimination group, and the worst of the low normal group, is arbitrary. The decision as to which group these borderline subjects belonged may perhaps be thought to have been intuitive on our part; but in the main it was based on various items in the composite picture, such as: 1. scores achieved on tests which yield quantitative scores, 2. types of error made (e.g., whether a digit on certain polychromatic tests was not seen at all or was merely read incorrectly, sometimes due to carelessness), 3. comparative scores and types of error made on repetition of tests, 4. performance with color threshold, hue difference and color matching (color aptitude) tests, as well as responses made to questioning, etc. For example, all of the subjects we have called "low normal" scored 75 or better on both the Ishihara (fifth edition) and the Rabkin test. The errors that were made were of the following nature: misreading a 13 for an 18; reading a "hidden digit," frequently incorrectly; or failing to see one of the digits of low visibility. In no case was more than one error made that could be considered of consequence. Further, in the cases where these slight errors were made, responses to other important tests, such as the 100-hue, dichotomy, anomaloscope, our Polychromatic Plates, etc. were of the order considered within normal limits either by the author of the test or by us as the result of experience.

In persons judged to have generally low color discrimination, the picture was quite different. All scored less than 75 on both the Ishihara and Rabkin tests, the errors made were of a more serious nature, and the responses to certain of the other tests, particularly the 100-hue test, were not within the limits of normality established for the test.

Low intelligence and lack of interest and concentration obviously contributed to a poor performance on certain tests, particularly for some individuals of the low normal group. When such factors were present in a color defective person, the performance score was unquestionably lowered, but the presence of a real color defect was evident in the type of responses given.

We feel confident that the subjects whose responses are used to evaluate the Rabkin test in this report and the Ishihara test in former reports,[1,2] are correctly classified in accordance with the scheme already outlined. We hope later to

TABLE I. Analysis of individual plates. Percentage of each type of color defective subject and of the low normals passing each plate of the Rabkin test (second edition).

Type of color defective subject	Number of cases	Plate No. 3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
		Percentage of each type passing individual plates																	
Anomalous Trichromat		%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%
Deuteranomalous	23	13	39	43	4	4	0	30	9	100	70	61	0	4	22	17	87	96	100
Protanomalous	12	8	92	0	0	8	0	0	0	75	33	0	0	0	8	0	83	92	100
Dichromat																			
Deuteranope	15	0	0	27	0	0	0	0	0	100	60	7	0	0	0	0	7	100	100
Protanope	17	6	100	0	0	0	0	0	0	59	0	0	0	29	0	0	0	82	100
Low Discrimination	7	14	86	0	29	29	0	29	14	100	72	86	0	14	0	57	86	100	100
All Color Defectives	74	8	58	19	4	5	0	12	4	86	46	28	0	9	8	11	50	93	100
Low Normals	32	94	100	78	100	94	44	100	91	100	97	100	91	100	100	94	100	97	100

report typical individual performances exemplifying these types.

The subjects used in the present report, then, included the following:

Anomalous Trichromats	
Deuteranomalous	23
Protanomalous	12
Dichromats	
Deuteranopes	15
Protanopes	17
Low discrimination	7
All color defectives	74
Low Normals	32

For all work with pigment test material a close approximation to I.C.I. Illuminant *C* was used. The illumination on the test material was about 25 ft.-c. Testing distance in case of polychromatic test plates was about 30 inches. Responses should be immediate (within 2 sec.). Hesitant or studied responses are viewed with suspicion. It should be mentioned also that in some of the Rabkin plates glossy pigments have been used. Special care must be taken, therefore, in illuminating these plates and in the viewing angle to avoid specular reflection which often affords a clue to either the correct or the incorrect response.

The results for the Rabkin test are summarized in Tables I–IV. They will be discussed under 4 headings: 1. Analysis of the individual plates; 2. Significance of performance scores on the entire test; 3. The differential classification of deuteranopia or protanopia and deuteranomaly or protanomaly; and 4. The differential classification of anomalous trichromasy and dichromasy.

1. Analysis of the Individual Plates

In Table I is shown for each type of color defective and for the low normals the percentage of cases who pass the individual Rabkin plates, designated by plate number. The first 5 horizontal rows of figures show these percentages for each type of color defective and demonstrate the value of the individual plates for detecting each type. The bottom two rows present a comparison between the percentages for the color defectives as a group and the low normals as a group. This comparison shows the value of each plate as a diagnostic or screening medium. In this table the following may be pointed out:

1. Plates 6, 7, 8, 9, 10, 14, 16, and 17 are not passed by any dichromats. (Of the plates which according to Rabkin are always failed by dichromats, Plates 3, 4, 5, 11, and 18 are not included in the above list.)[15] Plates 8 and 14 are also not passed by any anomalous trichromats or by any of those having low color discrimination; that is, these plates are not passed by any of our subjects having defective color vision. Plate 8, however, on which a faint reddish "5" is presented on a background of green lines, is passed by only 44 percent of the low normals, but Plate 14, which normal subjects read as "30" and dichromats as either "16" or "106," is passed by 91 percent of the low normals. Plate 14 is, therefore, the best single screening test of the series.

2. Plates 4 and 11 are singled out for special

[15] Rabkin states (page 37): "The fundamental symptom, permitting to differentiate the anomalous trichromasy from dichromasy, is the correct reading of one or several of the plates 3, 4, 5, 6, 11, 16, 17, and 18 by the anomalous trichromats. Other plates are read by them in the same way as by dichromats."

TABLE II. Significance of performance score, showing for each type of color defective subject and for the low normals (a) the average performance score achieved by each group; (b) the median or middle score; and (c) the range of scores from lowest to highest achieved by the individuals within each group. *Performance score is percentage of the 18 plates to which perfect responses were given.*

Type of color defective subject	Number of cases	Average score	Median score	Range of scores
Anomalous Trichromat				
Deuteranomalous	23	38	39	17–72
Protanomalous	12	28	25	17–39
Dichromat				
Deuteranope	15	22	22	17–28
Protanope	17	21	22	11–33
Low Discrimination	7	45	50	17–72
All Color Defectives	74	30	25	11–72
Low Normals	32	93	94	78–100

discussion because we are not able to confirm Rabkin's report as to the response of the color defective subject to these plates.

In Plate 4 a triangle composed of red disks is presented on a background of green and blue disks. Superimposed on the triangle is a circle composed in part of red and in part of green disks. Normal subjects respond "triangle." According to Rabkin, both protanopes and deuteranopes respond "circle." Among our observers all the deuteranopes and 61 percent of the deuteranomalous subjects reported "circle" (i.e., failed). All the protanopes and 92 percent of the protanomalous subjects, however, reported "triangle" (i.e., passed). They described the figure as a dark triangle on a lighter background. We believe that the brightness relationship between figure and background are so adjusted as to cause this plate to be passed by the protanopic and protanomalous types of subject.

Plate 11 is in construction similar to Plate 4, but in reverse color combination; that is, a triangle composed of green disks is presented on a background of red, orange and purple disks. Normal subjects respond "triangle." According to Rabkin, both protanopes and deuteranopes respond "circle." Among our observers all the deuteranomalous and deuteranopic subjects reported "triangle,"—the normal response; so also did 75 percent of the protanomalous subjects and 59 percent of the protanopes. Of the remaining protanomalous subjects only 8 percent reported "circle" and 17 percent failed to see any pattern. Of the remaining protanopes, only 18 percent reported "circle" and 23 percent failed to see any pattern. This plate, therefore, has no value as a test for red-green defect.

3. Plate 18 is the only one which shows a significant difference in the percentage of dichromats and anomalous trichromats who pass. Of the 32 dichromats only one (3 percent) passed this plate, but of the 35 anomalous trichromats, 30 (86 percent) passed. The possibilities of this plate for classification as to protanopia and deuteranopia are discussed in connection with Table IV.

4. Plates 19 and 20, which were intended for the detection of the tritanopic type of color defect (yellow-blue blindness) were passed by most of the color defectives tested. Plate 19 was passed by fewer of the protanopic type than of the deuteranopic type, and Plate 20 was passed by all of both types. We have been able to test too few subjects having the tritanopic type of defect to give an opinion on the value of these plates for detecting such cases.

2. Significance of Performance Scores on Entire Test

Table II shows the performance scores obtained on the entire Rabkin test. In this table are given for each type of color defective and for the low normals (a) the average score achieved by each group; (b) the median or middle score for each group; and (c) the range of scores from lowest to highest achieved by the individuals within each group. A score of 100 indicates that correct responses were given to all 18 significant plates; a score of 0 that no correct responses were given; and scores between 0 and 100 indicate the percentage of the 18 plates to which correct responses were given. As in Table I, the first 5 horizontal rows of figures give the data for each type of color defective, and the bottom 2 rows for the color defectives as a group and the low normals as a group.

In Table II it is seen that (a) the color defectives as a group have an average performance score of 30 and the low normals of 93; (b) deuteranomalous subjects have an average score of 38 and deuteranopes of 22; (c) protanomalous subjects

have an average score of 28 and protanopes of 21. On the basis of average performance score, then, the Rabkin test would seem to separate color defectives from color normals and deuteranomalous subjects from deuteranopes. However, for the purposes of screening and classification it is not the average score achieved by a group that is important, but it is the score achieved by the individual. This is shown in the last column of Table II where for each type of color defective is given the range of scores from lowest to highest achieved by the individuals having that type of defect. This column shows clearly the overlapping of scores between the various types of color defectives and thus demonstrates the impossibility of differentiating type of color defective on the basis of score alone. There is, however, no overlapping of scores between the group of color defectives taken as a whole and the group of low normals,—the lowest score of the low normal group being 78 and the highest score of the defective group being 72. In round numbers it would seem, then, that a critical score of 75 might be accepted as the dividing line between color defective performance and normal performance on the Rabkin test.[16]

The overlapping of scores among subjects exhibiting the various types of defective color vision is brought out still more clearly in Table III which shows the distribution of performance scores for each type. For representation in this table the following classes of performance scores were made: 0–6 (failure on all or all but 1 plate); 11–28 (passing 2–5 plates); 33–50 (passing 6–9 plates); 56–72 (passing 10–13 plates); 78–94 (passing 14–17 plates); and 100 (passing all plates). The overlapping of scores between the various types of color defectives is clearly shown in this table, and consequently the inability of the Rabkin test to yield a classification as to either type or extent of defect on the basis of performance score.

The following points are brought out in Tables I, II, and III:

1. The scores of anomalous trichromats and of subjects having generally low color discrimination sufficient to amount to a real defect have a wider range of scatter than those of dichromats (Table III). That is, as seems probable, their degree of defect may vary from a slight to an extreme anomaly or low discrimination, while dichromats having red-green defective vision present a more homogeneous group, at least when studied by means of tests designed primarily to detect red-green defective vision.

2. All color defectives of the red-green types, also cases having low color discrimination amounting to a color defect, are screened from those having normal and low normal color vision if a score of 75 is taken as the critical score in the Rabkin test (Tables II and III). In the group tested, no color defective achieved a score of more than 72 and no low normal of less than 78.

3. No clean-cut separation as to either extent or type of defect can be based on performance scores. While in general the dichromats had lower scores than the anomalous trichromats, 30 percent of the deuteranomalous and 58 percent of the protanomalous subjects had scores between 11 and 28, the range which includes all but one of the dichromats tested (Table III). *Scores cannot, therefore, be used to indicate either the extent or the type of the color defect.*

4. Of the dichromats tested, all passed at least one of the 16 plates designed to detect red-green

TABLE III. Distribution of performance scores. Percentage of each type of color defective subject and of the low normals having different ranges of performance score on Rabkin test (second edition).

Type of color defective subject	Number of cases	Percentage of each group achieving scores of: 0–6 (Number of plates passed: 0–1)	11–28 (2–5)	33–50 (6–9)	56–72 (10–13)	78–94 (14–17)	100 (18)
Anomalous Trichromat		%	%	%	%	%	%
Deuteranomalous	23	0	30	57	13	0	0
Protanomalous	12	0	58	42	0	0	0
Dichromat							
Deuteranope	15	0	100	0	0	0	0
Protanope	17	0	94	6	0	0	0
Low Discrimination	7	0	14	43	43	0	0
All Color Defectives	74	0	62	30	8	0	0
Low Normals	32	0	0	0	0	81	19

[16] Among the color defective subjects, only 3 percent passed as many as 13 plates (score 72); 4 percent passed 10 plates (score 56), and 9 percent passed 9 plates (score 50). The remaining 84 percent scored below 50. Among the low normal subjects, only 3 percent passed as few as 14 plates (score 78); 9 percent passed 15 plates (score 83); 13 percent passed 16 plates (score 89); 56 percent passed 17 plates (score 94), and 19 percent passed all 18 plates (score 100).

deficiency (Table I). This is mainly due to the fact that none of the deuteranopes failed plate 11, and none of the protanopes failed plate 4. A score of 22 is both the median and the most frequent score for dichromats (Table II).

3. The Differential Classification of Deuteranopia or Protanopia and Deuteranomaly or Protanomaly

So far we have discussed performance scores on the Rabkin test both as a means of detecting the presence of a color defect and as a means of classifying the type of defect, and have shown that the score achieved is adequate to separate defective from normal color vision when the test is properly administered, but it is not adequate to make the equally important classification as to type of defect and extent of defect. It remains to discuss *type of response* to certain plates as a clue to this classification. Since this classification constitutes a unique and important feature of the Rabkin test, the individual plates designed to indicate the type of red-green deficiency, as well as the test as a whole, will be discussed from that point of view.

Rabkin uses different series of plates to classify type of red-green defect among dichromats and anomalous trichromats.

Dichromats

According to Rabkin, deuteranopes and protanopes give a differential response to Plates 12, 13, 14, 16, 17, and 18. His key to diagnosis (classification) on the basis of type of response is given below.

Plate number	Reading of plates		
	Normal trichromats	Deuteranopes	Protanopes
12	12	12	—
13	*ct*	*ct*	*c*
14	30	16	106
16	96	6	9
17	*tc*	*c*	*t*
18	9–16	1, 2, 4, 6, 8	3, 5, 7

In practice, however, the following are some of the difficulties that are encountered in interpreting responses actually given to these plates by color defective subjects:

Plate 12.—Of our deuteranopes, 60 percent read "12," 7 percent read the digit incorrectly and 33 percent failed to see the digit. Of our protanopes, 88 percent failed to see the digit and 12 percent read it incorrectly. We feel that this type of plate yields ambiguous results. (a) The correct response can indicate either normal or deuteranopic vision, while failure to see the digit may be due either to a protanopic type of defect or to too low a color sensitivity to perceive the pattern, (33 percent of deuteranopes failed to see any pattern as well as 28 percent of those having low discrimination); and (b) it is impossible to interpret an incorrect reading of the digit.

Plate 13.—Here again, according to Rabkin normal trichromats and deuteranopes give the same response. In this case both types see "circle and triangle," while protanopes see only "circle." Of our deuteranopes, however, only 7 percent reported "circle and triangle" (normal and deuteranopic response), 80 percent saw "triangle" (ambiguous) and 13 percent saw no pattern (ambiguous). Of our protanopes, all saw "circle" (typical response for a protanope). For this plate we are unable to confirm Rabkin's schema.

Plate 14.—According to Rabkin, normal trichromats read "30," deuteranopes "16," and protanopes "106." The "3" and the "1" of this plate form a transformation pattern, the "0" is intended to be a vanishing digit for deuteranopes, and the "6" is a hidden digit supposedly seen only by color defectives. Of our deuteranopes, only 60 percent read "16." The remaining cases failed to see either the "6" or the "1," or else gave incorrect readings that could not be interpreted. Of our protanopes, only 41 percent read "106," 24 percent read "16" (the deuteranopic response), and the remainder failed to see any digits or read them incorrectly. A greater variety of responses was given to this plate than to any other in the test. This plate, therefore, has slight, if any, value.

Plate 16.—According to Rabkin, normal trichromats read "96," deuteranopes "6" and protanopes "9." In evaluating the response given to this plate and Plate 17, we amplified Rabkin's procedure as follows: If the classifying digit was read correctly and the other digit incorrectly, the correct reading was accepted as furnishing the

clue to classification. For example, "56" was accepted as diagnostic of deuteranopia; "95" of protanopia. On this basis all our deuteranopes and 71 percent of our protanopes were correctly typed. The remaining protanopes either saw no digits at all or read the classifying digit incorrectly.

Plate 17.—According to Rabkin, normal trichromats see "triangle and circle," deuteranopes see "circle" and protanopes see "triangle." Of our deuteranopes 93 percent were correctly typed by this plate, and of our protanopes 82 percent. Those who were not typed either saw no pattern or read it incorrectly.

Plate 18.—According to Rabkin, normal trichromats see as one-colored, the horizontal arrays 9–16; deuteranopes, the vertical arrays 1, 2, 4, 6, and 8; and protanopes, the vertical arrays 3, 5, and 7. In evaluating our responses to this plate, we accepted as diagnostic of deuteranopia or protanopia the selection of any one of the designated vertical arrays so long as no horizontal arrays were also selected. Even with this favorable interpretation, only 40 percent of our deuteranopes, and 59 percent of our protanopes were correctly typed. Of the remaining deuteranopes, 20 percent saw only horizontal arrays as one-colored (the normal response); 27 percent reported as one-colored some horizontal arrays and some of the vertical arrays 1, 2, 4, 6, 8 (deuteranomalous response); and 13 percent some horizontal arrays and some of the vertical arrays 3, 5, 7 (protanomalous response). Of the remaining protanopes, 35 percent reported as one-colored some horizontal arrays and some of the vertical arrays 3, 5, 7 (protanomalous response); and 6 percent saw all the arrays both horizontal and vertical, as one-colored (ambiguous).

In order to give the most favorable interpretation to the value of the Rabkin test as a whole for differentiating between deuteranopic and protanopic dichromats, we selected Plates 16 and 17 as eliciting the fewest ambiguous responses and as "typing" the greatest number of cases correctly. For a successful diagnosis we used the following criteria: (1) A case was called classified as to type of defect by the plate in question when one pattern was read correctly and the other not seen at all or read incorrectly. (2) A case was called classified as to type of defect by the test as a whole when a correct typing was obtained on either Plate 16 or 17. By these criteria, all of our deuteranopes and 94 percent of our protanopes were correctly classified as to type of red-green defect. The one case who was not typed failed to see a pattern on either plate. It may be added that the only plate on Rabkin's list that might have typed him was Plate 12 on which again he saw no pattern. Since 33 percent of our deuteranopes also failed to see a pattern on Plate 12, this failure to respond could scarcely be called a clue to diagnosis.

Anomalous Trichromats

To differentiate the type of red-green defect among anomalous trichromats, Rabkin employs Plates 13, 16, and 17.

Plate 13.—This plate is inadequate for this detection among anomalous trichromats for the same reason as among dichromats. Of our deuteranomalous cases, 61 percent reported "circle and triangle" (the normal and the deuteranopic response, according to Rabkin); 13 percent saw "triangle" (an unclassified response); 13 percent saw "circle" (the protanopic response); and 13 percent failed to see a pattern. All our protanomalous cases, however, reported "triangle."

Plate 16.—By this plate 74 percent of our deuteranomalous and 83 percent of our protanomalous cases were correctly classified. The remainder read either both digits or neither correctly, that is, they gave ambiguous responses.

Plate 17.—By this plate 70 percent of our deuteranomalous and 92 percent of our protanomalous subjects were correctly classified. As in Plate 16, the remainder saw correctly either both patterns or neither (ambiguous responses).

Applying the two criteria for a successful diagnosis by the test as a whole which was discussed in the section on *Dichromats*, we find 83 percent of our deuteranomalous and all our protanomalous subjects were correctly classified, none was incorrectly classified and 17 percent of the deuteranomalous were not classified. Had *Plate 13* been included among the diagnostic plates, as was done by Rabkin, 13 percent of the deuteranomalous would still have been unclassified and 4 percent would have been incorrectly classified as protanomalous.

It should be added that Rabkin suggests one

TABLE IV. The differential classification of deuteranopia or protanopia and deuteranomaly or protanomaly. Analysis of the value of Plates 16 and 17 as a means of differentiating between deuteranopic and protanopic types of red-green defective subject.

Type of color defective subject	Number of cases	Percentage of cases correctly classified by				Percentage of cases incorrectly classified by				Percentage of cases not classified by			
		Plate		Both plates	Either plate	Plate		Both plates	Either plate	Plate		Both plates	Either plate
		16	17			16	17			16	17		
Anomalous Trichromat		%	%	%	%	%	%	%	%	%	%	%	%
Deuteranomalous	23	74	70	52	83	0	0	0	0	26	30	48	17
Protanomalous	12	83	92	75	100	0	0	0	0	17	8	25	0
Dichromat													
Deuteranope	15	100	93	93	100	0	0	0	0	0	7	7	0
Protanope	17	71	82	65	94	0	0	0	0	29	18	35	6

further clue to the detection of the type of red-green defect among those anomalous trichromats who read both patterns correctly—that they be asked to designate which pattern is "plainer" or "clearer." We have found this procedure most unsatisfactory because the figure supposed to be seen clearer by the deuteranomalous type is also seen more clearly by those of normal color vision. This designation can, then, hardly be called evidence of deuteranomaly.

The results obtained from Plates 16 and 17 and from the test as a whole, by the interpretation outlined, are summarized in Table IV.

Our conclusion is that when evaluated in the revised manner described, the Rabkin test affords an excellent but not entirely perfect medium for the differential classification of type of red-green defect not only among dichromats, but also among anomalous trichromats.

4. The Differential Classification of Anomalous Trichromasy and Dichromasy

Rabkin is the first, so far as we know, to claim to be able to differentiate between anomalous trichromasy and dichromasy by means of polychromatic plates. It is of interest, therefore, to compare the results obtained by his method with those obtained with the Nagel Anomaloscope on the basis of which this differentiation was originally proposed and on which the classification of our subjects is largely based.

Rabkin's method is summed up in his statement, page 37, as follows: "The fundamental symptom permitting to differentiate the anomalous trichromasy from dichromasy is the correct reading of one or several of the plates 3, 4, 5, 6, 11, 16, 17, and 18 by the anomalous trichromats. Other plates are read by them in the same way as by dichromats." He lays special emphasis on the reading of Plate 18. According to him, protanopes see as one-colored columns 3, 5, and 7; deuteranopes, columns 1, 2, 4, 6, and 8; anomalous trichromats with a high degree of defect approaching dichromasy see some horizontal and some vertical arrays as one-colored, and anomalous trichromats with a low degree of defect approaching normal trichromasy see only the horizontal arrays as one-colored. The defect in the latter group is detected by their failure to read correctly other plates of the test.

According to our results, Rabkin's method is not adequate to differentiate between anomalous trichromats and dichromats. All our anomalous cases were diagnosed as anomalous by his procedure in that they read correctly one or more of the plates listed above, failed several of the remaining plates, and in Plate 18 saw as one-colored either horizontal arrays alone (87 percent of the deuteranomalous and 83 percent of the protanomalous) or some horizontal and some vertical arrays (13 percent of the deuteranomalous and 17 percent of the protanomalous). These cases were, then, correctly classified by Rabkin's criteria.

However, by his first criterion, all of our dichromats are also classified as anomalous trichromats. It will be remembered from the section on "Analysis of the Individual Plates" that all our deuteranopes passed Plate 11 and all our protanopes passed Plate 4. Further, 33 percent of the deuteranopes and 41 percent of the protanopes passed one or more other plates in the critical

list. All our dichromats, then, read correctly at least one, and 38 percent read two or more of the specified plates. His second criterion (i.e., the response to Plate 18) is also inadequate to separate dichromats from anomalous trichromats. By it, only 40 percent of our deuteranopes and 59 percent of our protanopes are classified as dichromats. In brief, a total of only 50 percent of our dichromats are correctly classified as such by his second criterion and none are so classified by his first criterion.

COMMENTS

There are two ways in which the possibilities of the Rabkin test for the differential classification of type and degree of defect could be improved. (1) *In the pigments used.* Rabkin makes no use of the second neutral range in the color scheme of deuteranopes as contrasted with that of protanopes, i.e., in the red-purple region for the former and the red region for the latter. This would have been especially useful in Plate 18. The vertical arrays supposed to be seen one-colored by deuteranopes would then have consisted of squares of red-purple and green of the same brightness, instead of red and green; those supposed to be seen as one-colored by protanopes, of squares of lighter red and darker blue-green. The expectation that characteristic vertical arrays could be selected as one-colored would thus be increased and, since horizontal arrays would clearly appear to dichromats as squares of blue, yellow and gray, the expectation that horizontal arrays would be selected would be decreased. (2) *In the chroma of pigments used.* It is logical that lower chroma pigments are needed to detect type of defect in those whose defect is less pronounced, and higher chroma for this detection in extreme cases. Using similar type plates in a series of graded chromas would aid also in the determination of the degree or extent of the defect. We have made plates somewhat similar to Rabkin's Plate 18 involving both these features. On preliminary trial these plates seem quite adequate for the dual differentiation, i.e., the separation of anomalous trichromats from dichromats, and the classification within each group of type and degree of red-green defect. This work will be reported later.

CONCLUSIONS

The Rabkin test is the most informative of any of the pseudo-isochromatic plate tests now on the market. Used with care in administration, scoring, and interpretation, it yields much more information concerning the type of defect of the examinee than does, for example, the Ishihara test or the American Optical Company's selection of plates. As the result of our study of the test, changes in scoring certain of the plates are recommended, a revised procedure for classifying type of red-green defect is described and constructive comments are offered for the improvement of some of the plates. In addition the following general conclusions concerning the second edition of the Rabkin test seem justified.

1. Properly administered the test affords a good device for screening the color defective from the color normal if 75 is taken as the critical performance score.

2. No analysis as to type or extent of defect can be based on score alone.

3. The test taken as a whole affords an excellent means of classifying red-green dichromats into the two groups: protanopia and deuteranopia; also of classifying red-green anomalous trichromats into the two groups: deuteranomalous and protanomalous.

4. The test as now constructed is, however, not adequate to differentiate between anomalous trichromasy and dichromasy.

To clarify 3 and 4, the conclusions might be stated in this manner: If we designate both dichromats and anomalous trichromats as either "proto's" (predominantly red-defective) or "deutero's" (predominantly green-defective), and if we further designate anomalous trichromats as "ouses" and dichromats as "opes," then the Rabkin test very effectively aids us in differentiating the "proto's" from the "deutero's" but is very inefficient in separating the "ouses" from the "opes."

Absorption and Excitation of Zinc Silicate Phosphors

CHUNG KWAI LUI
Research Department, Westinghouse Electric Corporation, Bloomfield, New Jersey
(Received April 12, 1945)

Excitation spectrum of zinc silicate phosphor showed two bands. A major band is located between 2250A and 2900A, and a minor one between 1950A and 2250A. Absorption measurements of clear willemite crystals showed two similar bands located at about the same positions. Absorption measurements for zinc silicate phosphor showed a band located between 2000A and 3000A. These results indicate that the excitation spectrum of a phosphor also measures the absorption of that phosphor.

INTRODUCTION

IN studying fluorescence, the absorption characteristic of fluorescent materials is very essential. When the fluorescent materials are gases, liquids, solutions, or transparent crystals, the absorption spectra can easily be determined; but when the fluorescent material is a powder, the excitation method is generally used as a measurement of absorption. Therefore if absorption measurements can be obtained for powdered material the result should agree with the excitation measurements.

The absorption and excitation spectra of zinc silicate phosphor reported by F. A. Kröger[1] are shown in Fig. 1. His results showed that the excitation spectrum of one percent manganese-activated zinc silicate phosphor had a peak at about 2500A. However, the absorption spectrum while high at 2250A showed only one-sixth as much absorption at 2537A. It would appear that either the excitation measurement or the absorption measurement is in error. The range of excitation spectrum for zinc silicate phosphor given by Amick[2] is between 2200A and 2960A with a maximum at 2537A. Marden and Beese[3] showed a strong excitation for the zinc silicate phosphor at about 2400A. Since the excitation spectrum of zinc silicate phosphor as given by Kröger agrees very well with those of other investigators, the absorption measurement of Kröger may be in error. An improved method to measure the absorption of phosphors is presented in this paper.

[1] F. A. Kröger, Physica **6**, 764–778 (1939).
[2] C. L. Amick, *Fluorescent Lighting Manual* (McGraw-Hill Book Company, Inc., New York, 1942).
[3] J. W. Marden and N. C. Beese, Illum. Eng. **36**, 235–242 (1941).

APPARATUS

A grating spectrograph was used throughout this investigation. The slit, grating, and photographic film all lie on the same circumference of a circle one meter in diameter as described by Beese[4] except that vacuum is not used in this experiment. The ultraviolet light source was a high pressure mercury lamp of the H3 type.

EXPERIMENT

In order to get the true absorption of a solid substance, the light must pass through the substance. The best substance would be a clear thin crystal. Unfortunately no satisfactory method is available at present to grow the artificial crystals from fluorescent compounds. However, small clear willemite crystals about 0.5 mm wide and 3 mm long were obtained from the New Jersey Zinc Company. The crystal was carefully supported in front of the slit of the spectrograph with a piece of soft wax. The spectra were taken with varying exposure times on super-speed ortho portrait Eastman Kodak film. On the same film a mercury reference spectrum was also taken. The intensities of the mercury lines of the different spectra were measured with a Leeds and Northrup recording microphotometer. The decrease in intensities of the mercury lines at different wave-lengths measured the absorption of willemite crystal. Since the intensities of the mercury lines decreased toward the shorter wave-length ultraviolet, necessary corrections were made. The absorption spectrum at different wave-length as shown in Fig. 2 has two maxima. The major

[4] N. C. Beese, J. Opt. Soc. Am. **29**, 278–282 (1939).

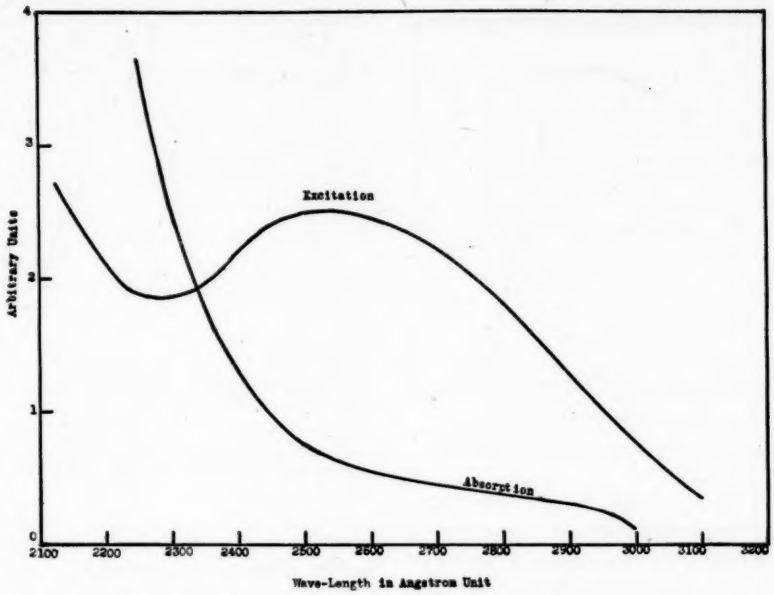

FIG. 1. Absorption and excitation of zinc silicate phosphor from Kröger's result.

band is located between 2170A and 2900A, with a peak at about 2600A and a minor one between 1950A and 2170A.

Since it is impossible to obtain clear crystals of fluorescent compounds at present, an attempt to measure the absorption of fluorescent powder was made.

A preliminary trial was made by coating a thin layer of fluorescent powder on a piece of quartz plate. The result in this case showed no characteristic absorption in the ultraviolet region. Examination of the specimen under a low power microscope showed numerous small holes. Therefore the light must have passed through the holes instead of passing through the particles of the phosphors. In order to minimize the number of holes, a liquid medium might be introduced so that a uniform continuous thin film of fluorescent powder is obtained. Stutz[5] used castor oil as the dispersing medium to measure the absorption of paint pigments. Castor oil as shown by his paper has characteristic absorption below 2537A. Since the excitation spectrum of zinc silicate phosphor has a peak at about 2537A, its absorption spectrum may be at about the same region. Therefore castor oil is not suitable to use in this case. Glycerin, however, has been found quite satisfactory.

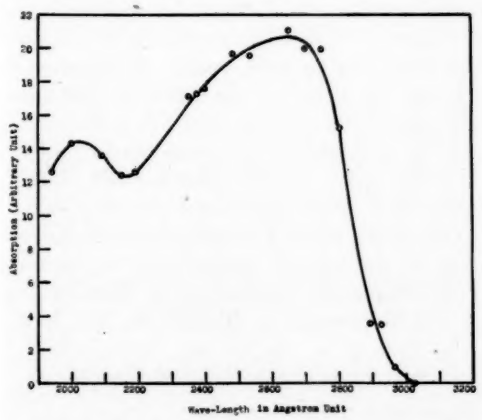

FIG. 2. Absorption spectrum of willemite crystal.

The best way to prepare the specimen is to make a paste of the fine powder with a few drops of glycerin and carefully spread it between two quartz plates. The dimensions of the quartz plate are 1 mm thick and 2 inches square. The sample is examined with a low power microscope to see that the film is continuous over a large area.

[5] G. F. A. Stutz, J. Frank. Inst. **202**, 89–98 (1926).

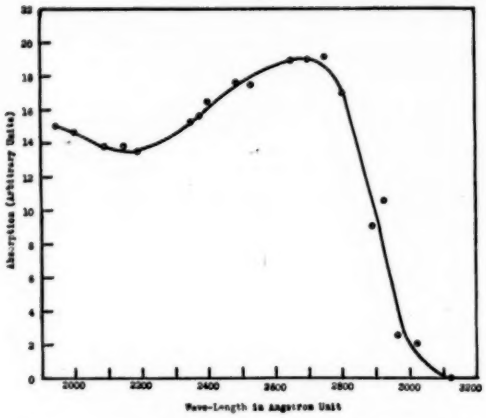

FIG. 3. Absorption spectrum of zinc silicate phosphor.

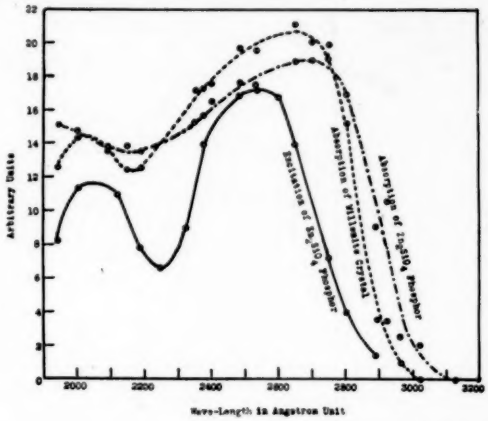

FIG. 4. Absorption of zinc silicate phosphor and willemite crystal compared with excitation of zinc silicate phosphor.

The specimen is interposed between the light source and slit of the spectrograph. Spectrograms are then taken with different exposure times. On the same film a comparison spectrum is taken with light only passing through a film of glycerin about the same thickness spread between two quartz plates. The method of measuring the absorption is the same as described for willemite crystals. It was thus possible by this method to obtain a relative measure of the true absorption of the fluorescent powder. The absorption spectrum for zinc silicate phosphor is shown in Fig. 3. In order to make sure that there is no characteristic absorption due to either glycerin or quartz in the ultraviolet wave-length range, that is, between 2000A and 4000A, spectrograms were taken of the light source alone and then with light passing through the combination of glycerin and quartz plates. No noticeable characteristic absorption was detected. The absorption curve for the zinc silicate powder showed a maximum at about 2700A.

For comparing the absorption spectra of willemite crystals, and zinc silicate phosphor with the excitation spectrum, curves were plotted on the same coordinate as shown in Fig. 4. The excitation spectrum was measured by the method described by Marden and Beese[3] except that the grating spectrograph was used instead of the quartz prism spectrograph.

DISCUSSION

The absorption spectrum for a clear willemite crystal as shown in Fig. 4 agrees very well with the excitation spectrum. The slight shift of the peak may be due to the non-uniformity of the energy of the light source. The absorption spectrum of zinc silicate phosphor deviated from the excitation spectrum in the short ultraviolet. This anomaly may be due to the scattering phenomena at such short wave-lengths. However, the peak of absorption is located at about 2700A, which is closer to the excitation spectrum. Better results of absorption measurement of fluorescent powder may be obtained if an intense light source of uniform energy throughout the range of wave-lengths from 2000A to 4000A were available. From the results of this experiment it may be concluded that measurement of the excitation spectrum of a phosphor will also give a measure of the absorption of that phosphor.

The author wishes to express her deep appreciation for the helpful suggestions from members of the Research Department of Westinghouse Electric Corporation at Bloomfield, New Jersey.

Letter to the Editor

A Simple Formula for Lens Design

JOHN R. MILES
Optical Engineering Department, Eclipse-Pioneer Division, Bendix Aviation Corporation, Teterboro, New Jersey
March 15, 1945

A FORMULA is presented which may be of interest to those engaged in designing lenses. The writer's derivation of the expression is simple mathematically, but has proved to be useful.

In designing lenses with a minimum over-all thickness, a large relative aperture, and a moderate field of about 10°, the best choice of glasses is essential. The writer noted that the actual Petzval curvature could not be directly found by any known formula for a given combination of glasses. This meant that the combination which would have the acceptable radius of Petzval field curvature and also have the greatest difference of V could not be directly chosen. Others had set a permissible difference of V and then taken the combination giving the lowest difference of V/N. The V referred to is $(N_D-1)/(N_F-N_C)$ and the N's are indices of refraction in D, F, and C lights respectively.

The writer then made the following original derivation for the purpose of obtaining, from the available glasses, a combination of glasses having a low enough Petzval field curvature consistent with basic color correction and having the best possible V difference:

$$\text{Petzval Sum} \equiv P\Sigma = \frac{1}{f_A N_A} + \frac{1}{f_B N_B}, \tag{1}$$[1]

and for basic color correction, the focal lengths of the crown and flint portions are respectively,

$$\left.\begin{aligned} f_A &= F\left(\frac{V_A - V_B}{V_A}\right) \\ f_B &= F\left(\frac{V_B - V_A}{V_B}\right) \end{aligned}\right\} \begin{array}{l}\text{for thin, closely spaced lenses} \\ \text{where } F \text{ is focal length of} \\ \text{complete system.}\end{array} \tag{2}$$[2]

Then substituting Eqs. (2) into (1),

$$P\Sigma = \frac{1}{F(V_A - V_B)(N_A/V_A)} + \frac{1}{F(V_B - V_A)(N_B/V_B)}, \tag{3}$$

or

$$P\Sigma = \frac{1}{F}\,\frac{1}{V_A - V_B}\left(\frac{V_A}{N_A} - \frac{V_B}{N_B}\right) \equiv \frac{1}{\rho}. \tag{4}$$

Hence for any given focal length F, the Petzval field radius ρ corresponds with

$$MR = \frac{V_A - V_B}{\left(\frac{V_A}{N_A} - \frac{V_B}{N_B}\right)}, \tag{5}$$

for use with a basic glass MR table which is independent of focal lengths and hence generally applicable to all focal lengths,

$$\rho = F \cdot MR, \tag{6}$$

as is convenient for determining the Petzval radius ρ once the focal length is known.

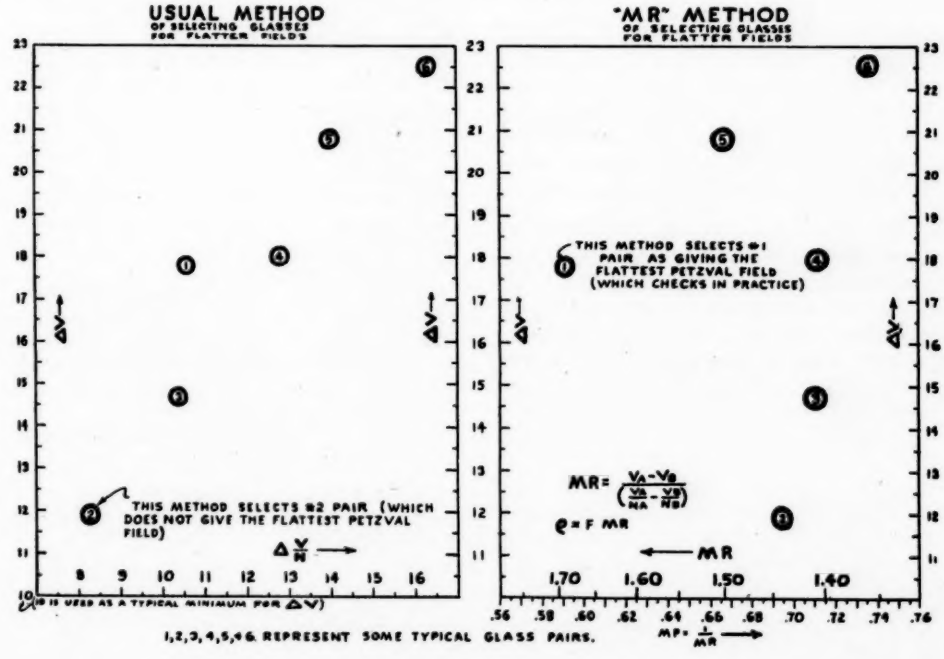
USUAL METHOD
OF SELECTING GLASSES
FOR FLATTER FIELDS
THIS METHOD SELECTS #2 PAIR (WHICH DOES NOT GIVE THE FLATTEST PETZVAL FIELD)
ΔV
"MR" METHOD
OF SELECTING GLASSES
FOR FLATTER FIELDS
THIS METHOD SELECTS #1 PAIR AS GIVING THE FLATTEST PETZVAL FIELD (WHICH CHECKS IN PRACTICE)
MR = (VA - VB) / (VA/NA - VB/NB)
ρ = F MR
MR
1.70
1.60
1.50
1.40
1,2,3,4,5,6 REPRESENT SOME TYPICAL GLASS PAIRS.

FIG. 1.

This does not involve the assumption that

$$\frac{1}{f_A N_A}+\frac{1}{f_B N_B}=0, \tag{7}$$

heretofore made by others, which is not true for any useful pair of glasses known.

The value MR of condition (5), for a given combined focal length, corresponds with the radius of curvature of the Petzval field. Consequently the value MR may be used directly to select that pair of available useful glasses which gives a satisfactory Petzval field without loss of color correction and with the lowest other aberrations.

Frequently in the past, the designer, in following the usual method as illustrated on the left side of the figure, would adopt a pair of glasses such as (2) only because it most nearly satisfies the Petzval condition.

However, by the MR method as demonstrated on the right-hand graph of the figure, he can confidently select the best pair[1] as to field curvature, while by the usual method he would get a relatively poor result.

Although this illustration of methods shows only six pairs of glasses, it is obvious that all available pairs could be plotted and limits could be set, for example, at $MR>1.5$ and at $\Delta V>10$ so that a large selection could be had either for greater ΔV or for improving secondary spectrum.

This procedure is more useful now that small air-spaces can be more generally used to lower the V-difference usable, since low-reflecting films reduce otherwise-excessive loss of light or the otherwise-possible presence of "ghost" reflections at small air-spaces.

[1] Conrady, *Applied Optics and Optical Design*, Part one (1929), Page 289.
[2] Conrady, *Applied Optics and Optical Design*, Part one (1929), Page 149.

A Message to All Interested in Physics

From the Chairman of the Building Fund Committee American Institute of Physics

YOU have probably heard the story of the young mother of the baby that was a year old before the last payment was made to the doctor. Her exclamation was, "Now we own the baby." The headquarters building has been occupied for over a year but we don't yet "own the baby."

The only thing that has been asked of physicists and friends of physics is that they give a firm, decisive "Yes or No answer" to the question, "Will you help in the purchase of the Physics Headquarters Building?" and that a "Yes" answer be accompanied by a pledge or contribution. No firm of experts has been hired. No high powered methods have been used. A committee of physicists has done the work.

It has been hoped that practically all physicists and many friends of physics would make a contribution even though small. In some departments of Physics in universities there has been a 100 percent response, with graduate assistants necessarily giving small contributions. A large number of contributions of between one and ten dollars are needed as well as some more of the larger contributions from those who "have not yet got 'round to it."

The amount necessary for the purchase of the property and essential repairs was $81,000, of which $6,000 remains to be subscribed by *individual physicists and friends of physics*. The amount needed for adapting the building to present needs and furnishing it was $30,000, of which $4,000 remains to be contributed by industrial organizations interested in physics.

Frankly, I am proud of the fact that physicists and friends of physics have already raised over $100,000 to help promote the science of physics and thus show our interest and our faith in the future of our field. But I am disappointed in lack of response on the part of a large number. If you are one of these, I appeal to you to send a contribution, no matter how small, to the *Headquarters Building Fund Committee*, care of the *American Institute of Physics, 57 East 55 Street, New York City*, or to me at *Norwich University, Northfield, Vermont*.

If, by any chance, you are not familiar with the project it is suggested that you refer to the Review of Scientific Instruments, August 1943, page 241; and the Journal of Applied Physics, September 1943, pages xi and xii, and October 1943, pages 499–509.

HOMER L. DODGE
Chairman, Building Fund Committee